目　　录

第一章 矩阵理论

在本章中我们将介绍“多元统计分析”课程中所需要矩阵的有关理论，凡是在通常线性代数教材中已有的材料我们将不加证明或者略加说明地进行引用。虽然本章的内容并非本教材的中心内容，但其有关的结论对于以后各章的学习都是十分重要的。在本章中我们引进了在通常线性代数教材中很少提及的一些矩阵运算，如拉长向量 vec$A$ 将帮助我们把随机矩阵的讨论化为熟知的随机向量的讨论。而 Kronecker 乘积 $A\otimes B$ 在处理随机矩阵的协方差矩阵时有着十分简明的意义。不同的 Jacobian 行列式的导出对于求出各种随机矩阵相应的分布密度是至关重要的，希望读者能熟练掌握。

§1.1 矩阵的有关定义及其运算

（一） 矩阵概念及其运算

令 R 是由实数组成的域，在 R 中任一组数字的四则运算总能进行。设 $a_{11},\cdots,a_{pq}$ 是 R 中 pq 个元素，称由这些元素组成的 p 行 q 列矩形方块

$$\begin{pmatrix} a_{11} & a_{12} \cdots a_{1q} \\ a_{21} & a_{22} \cdots a_{2q} \\ & \cdots\cdots \\ a_{p1} & a_{p2} \cdots a_{pq} \end{pmatrix}$$

为一个 $p\times q$ 阶**矩阵**，通常记为大写字母 $A=(a_{ij})_{p\times q}$，其中 a_{ij} 是 A 的第 i 行第 j 列上的元素。如果 A 的诸元素都为 0，则称 A 是一个 $p\times q$ 阶**零矩阵**，记为 $O_{p\times q}$，在不致误解时可简记为 O。如果 $p=q$，称 A 为 p 阶**方阵**。当 $p=1$ 时也称 A 为 q 维**行向量**，当 $q=1$ 时也称 A 为 p 维**列向量**。

以后总是用

$$\alpha=\begin{pmatrix}a_1\\a_2\\\vdots\\a_p\end{pmatrix}$$

表示由元素 $a_1,\cdots,a_p$ 组成的列向量，而用 $\alpha'=(a_1,a_2,\cdots,a_p)$ 表示由元素 $a_1,a_2,\cdots,a_p$ 组成的行向量。

如果 p 阶方阵中所有非对角线上元素全为 0 (即 $a_{ij}=0, i\neq j$)，则称 A 为 p 阶**对角阵**。对角线元素为 $\lambda_1,\cdots,\lambda_p$ 的对角阵记为 Λ 或者 $\mathrm{diag}(\lambda_1,\cdots,\lambda_p)$。特别当 λ_i 全为 1 时，相应的对角阵称为 p 阶**单位方阵**，可简记为 I_p。

称在主对角线下方诸元素全为 0 的方阵为**上三角矩阵**，称在主对角线上方诸元素全为 0 的方阵为**下三角矩阵**。

称将矩阵 A 的诸行（列）换成相应的诸列（行）所得到的新矩阵为 A 的**转置矩阵**，记为 A'。从而若 $A=(a_{ij})_{p\times q}$，则 $A'=(a_{ji})_{q\times p}$。

如果方阵 A 满足 $A=A'$，称 A 为**对称矩阵**(或对称阵)。

有时我们把一个 $p\times q$ 阶矩阵 A 表示成分块形式

$$A=\begin{pmatrix}A_{11}&\cdots&A_{1n}\\\vdots&&\vdots\\A_{m1}&\cdots&A_{mn}\end{pmatrix}$$

其中 A_{ij} 是 A 的一个 $p_i\times q_j$ 阶子矩阵 ($i=1,\cdots,m, j=1,\cdots,n$) 且 $\sum_{i=1}^{m}p_i=p$，$\sum_{j=1}^{n}q_j=q$。

下面介绍矩阵的运算。

两矩阵 A,B (具有相同的列数和行数)的和规定为

$$A+B=(a_{ij})+(b_{ij})=(a_{ij}+b_{ij})$$

一实数 λ 与矩阵 A 的乘积规定为

$$\lambda\cdot A=A\cdot\lambda=(\lambda a_{ij})$$

如果矩阵 A 的列数等于 B 的行数，即 $A=(a_{ij})_{p\times q}$，$B=(b_{jk})_{q\times r}$，则规定 A 与 B 的积为

$$A\cdot B=\left(\sum_{j=1}^{q}a_{ij}b_{jk}\right)_{p\times r}$$

即 $A\cdot B$ 是一个 $p\times r$ 阶矩阵，它的第 (i,k) 元素是 $\sum_{j=1}^{q}a_{ij}b_{jk}$。

可以验证上述运算具有以下性质：

1) $A+B=B+A$。

2) $(A+B)+C=A+(B+C)$。

3) $A+(-1)A=O$。

4) $(\lambda+\mu)A=\lambda A+\mu A$。

5) $\lambda(A+B)=\lambda A+\lambda B$。

6) $\lambda(\mu A)=(\lambda\mu)A$。

7) $(A+B)C=AC+BC$。

8) $(AB)C=A(BC)$。

9) $I\cdot A=A\cdot I=A$。

若 A 和 B 为同阶上(下)三角矩阵，那么它们的和 $A+B$ 以及它们的积 $A\cdot B$ 也是同类型的上(下)三角矩阵。

转置矩阵具有下列性质：

1) $(A')'=A$。

2) $(A+B)'=A'+B'$。

3) $(A\cdot B)'=B'\cdot A'$。

若矩阵 A 和 B 具有相同的行数与列数且按同样方式分块，即 $A=(A_{ij})$，$B=(B_{ij})$，那么

$$A+B=(A_{ij}+B_{ij}).$$

若 $m\times n$ 阶矩阵 A，$n\times p$ 阶矩阵 B 分别划分为 $A=(A_{ij})$，$B=(B_{jk})$，其中 A_{ij} 为 $m_i\times n_j$ 阶矩阵，B_{jk} 为 $n_j\times p_k$ 阶矩阵，且 $\sum_i m_i=m$，$\sum_j n_j=n$，$\sum_k p_k=p$。那么

$$AB = \left(\sum_j A_{ij}B_{jk}\right)$$

（二） 矩阵的直积

如果 A 是 $m\times n$ 阶矩阵，B 是 $p\times q$ 阶矩阵，它们的**直积**（或者称 **Kronecker** 积）是如下定义的 $mp\times nq$ 阶矩阵

$$A\otimes B=(a_{ij}B)$$

其中 $A=(a_{ij})$。它表示以 $a_{ij}B$ 为它的第 (i,j)块的分块矩阵.

直积运算具有以下性质:

1) $0\otimes A=A\otimes 0=O$。

2) $(A_1+A_2)\otimes B=A_1\otimes B+A_2\otimes B$。

3) $A\otimes(B_1+B_2)=A\otimes B_1+A\otimes B_2$。

4) $\alpha A\otimes\beta B=\alpha\beta\cdot A\otimes B$。

5) $A_1A_2\otimes B_1B_2=(A_1\otimes B_1)(A_2\otimes B_2)$。

6) $(A\otimes B)'=A'\otimes B'$。

证明　我们仅证明 5) 式，其余请读者自行完成。

设 $A_1=(a_{ij})_{p\times m}$，$A_2=(c_{jk})_{m\times n}$，B_1 是 $s\times t$ 阶矩阵，B_2 是 $t\times u$ 阶矩阵。于是

$$A_1A_2\otimes B_1B_2=\left(\sum_{j=1}^m a_{ij}c_{jk}\right)_{p\times n}\otimes B_1B_2=\left(\sum_{j=1}^m a_{ij}c_{jk}B_1B_2\right)$$

是一个 $ps\times nu$ 阶矩阵，它的第 (i,k) 块是 $\sum_{j=1}^m a_{ij}c_{jk}B_1B_2$。而

$$A_1\otimes B_1=(a_{ij}B_1),\quad A_2\otimes B_2=(c_{jk}B_2)$$

按分块矩阵乘法有

$$(A_1\otimes B_1)(A_2\otimes B_2)=\left(\sum_{j=1}^m a_{ij}c_{jk}B_1B_2\right)$$

从而 5) 式成立。

令 $\mathrm{vec}A$ 表示把 $p\times q$ 阶矩阵的诸行向量依次排列成一个拉长的行向量再加以转置而得到的 pq 维列向量，即

$$\text{vec}A = (a_{11}, a_{12}, \cdots, a_{1q}, a_{21}, a_{22}, \cdots, a_{2q}, \cdots, a_{p1}, \cdots, a_{pq})'$$

通过这种记号可以把有关矩阵的运算化为向量的运算。这时，对于矩阵方程

$$AXB = C \tag{1.1}$$

其中 A, X, B, C 分别为 $m \times n$ 阶，$n \times r$ 阶，$r \times q$ 阶，$m \times q$ 阶矩阵，利用直积记号可以改写成另一等价形式：

$$(A \otimes B')\text{vec}X = \text{vec}C \tag{1.2}$$

对此只须注意到把(1.1)和(1.2)具体写成线性方程组后都是

$$\sum_{j=1}^{n} \sum_{k=1}^{r} a_{ij} x_{jk} b_{kl} = c_{il} \qquad (i = 1, \cdots, m, l = 1, \cdots, q).$$

由此可得

定理 1.1 $\text{vec}(AXB) = (A \otimes B')\text{vec}X$.

例如，对于向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 有

$$\text{vec}(\boldsymbol{x} \cdot \boldsymbol{y}') = \boldsymbol{x} \otimes \boldsymbol{y}.$$

（三） 向量的相关和 Schmidt 正交化方法

为了对矩阵有更深刻的理解，下面我们复习一些向量空间（线性空间）的知识。

设 R^p 表示实数域 R 上的全部 p 维向量，那么任一 p 维向量可以表示为 R^p 空间上的一个点，它作为特殊矩阵，可以在 R^p 上定义加法和数乘运算。显然 R^p 关于这些运算是封闭的，我们称 R^p 是一个**向量空间**。

对于 $p \times q$ 阶矩阵 $A \doteq (a_{ij})$，我们记它的列向量为 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_q$，行向量为 $\boldsymbol{a}'_{(1)}, \cdots, \boldsymbol{a}'_{(p)}$，那么 A 可以写成

$$A = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_q) \text{ 或者 } A = (\boldsymbol{a}_{(1)}, \cdots, \boldsymbol{a}_{(p)})' \tag{1.3}$$

从而矩阵 A 既可看成 R^p 空间中的 q 个点，又可看成 R^q 空间中的 p 个点。另外由于矩阵 A 与 $\text{vec}A$ 是一一对应的，它又可看成是 R^{pq} 空间中的一个点。

对于一组向量 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_q$，如果存在不全为 0 的一组实数

$c_1,\cdots,c_q$，使得

$$c_1\boldsymbol{a}_1+c_2\boldsymbol{a}_2+\cdots+c_q\boldsymbol{a}_q=0$$

成立，则我们称向量组 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q$ 是**线性相关**的。否则称向量组 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q$ 是**线性无关**的。

容易证明下列性质：

1）任何包含零向量的向量组总是线性相关的。

2）设 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q$ 是非零向量，则它们是线性相关的充分必要条件是：一定存在某一 i 使得

$$\boldsymbol{a}_i=\sum_{j\neq i}c_{ij}\boldsymbol{a}_j.$$

成立，其中 c_{ij} 不全为 0（$j=1,\cdots,q,\ j\neq i$）。即存在 i，使得 $\boldsymbol{a}_i$ 是其他向量 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_{i-1},\boldsymbol{a}_{i+1},\cdots,\boldsymbol{a}_q$ 的线性组合。

若向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 满足 $\boldsymbol{a}'\boldsymbol{b}=0$，则我们称 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ **正交**。

3）若非零向量 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q$ 两两正交，则它们必线性无关。事实上，若存在不全为 0 的实数 c_i 使得 $\sum_{i=1}^{q}c_i\boldsymbol{a}_i=0$，则由

$$\left(\sum_{i=1}^{q}c_i\boldsymbol{a}_i\right)'\left(\sum_{i=1}^{q}c_i\boldsymbol{a}_i\right)=\sum_{i=1}^{q}c_i^2\boldsymbol{a}_i'\boldsymbol{a}_i=0$$

推出 c_i 全为 0，得出矛盾。

记 $\|\boldsymbol{a}\|=\sqrt{\boldsymbol{a}'\boldsymbol{a}}$，它表示向量 $\boldsymbol{a}$ 的**长度**，称长度为 1 的向量为**单位向量**。

给出 R^p 中一组向量 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q$，考虑由这些向量的所有可能的线性组合 $\sum_{i=1}^{q}c_i\boldsymbol{a}_i$ 组成的集合

$$\left\{\sum_{i=1}^{q}c_i\boldsymbol{a}_i,\ c_1,\cdots,\ c_q\ \text{是实数}\right\}$$

它显然对于向量加法和数乘运算是封闭的，从而构成 R^p 中的一

个**线性子空间**，称为由向量 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_q$ 张成的子空间。记为 $\mathscr{L}(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q)$。对于 $p\times q$ 阶矩阵 A，按(1.3)，以后将使用记号

$$\mathscr{L}(A)=\mathscr{L}(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q),\mathscr{L}(A')=\mathscr{L}(\boldsymbol{a}_{(1)},\cdots,\boldsymbol{a}_{(p)})$$

记 $\mathscr{L}$ 是 R^p 的一个线性子空间，如果存在一组线性无关的向量 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q$，使得 $\mathscr{L}=\mathscr{L}(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q)$，则称 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q$ 是线性子空间 $\mathscr{L}$ 的一组**基底**，q 称为 $\mathscr{L}$ 的维数，这时 $\mathscr{L}$ 中的任一向量都可以通过它的基底向量的线性组合来表示。如果这组基底还满足

$$\boldsymbol{a}_i'\boldsymbol{a}_j=0\ (i\neq j),\ \boldsymbol{a}_i'\boldsymbol{a}_i=1\quad(i、j=1,\cdots,q)$$

则 $\{\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q\}$ 又称为$\mathscr{L}$的**标准正交基底**。

例如 R^3 空间中，向量 $(1,0,0)'$，$(0,1,0)'$，$(0,0,1)'$ 构成了 R^3 的标准正交基底。

如果 $\mathscr{L}(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q)$ 中 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q$ 线性无关，我们可以如下构造一组新的标准正交基底。

令

$$\boldsymbol{b}_1=\boldsymbol{a}_1,\ \boldsymbol{b}_2=\boldsymbol{a}_2-h_{21}\boldsymbol{b}_1\ (h_{21}=\boldsymbol{a}_2'\boldsymbol{b}_1/\boldsymbol{b}_1'\boldsymbol{b}_1)$$

那么显然有 $\boldsymbol{b}_1'\boldsymbol{b}_2=0$。依次再令

$$\boldsymbol{b}_k=\boldsymbol{a}_k-h_{k1}\boldsymbol{b}_1\cdots-h_{k,k-1}\boldsymbol{b}_{k-1}$$

$$(h_{ki}=\boldsymbol{a}_k'\boldsymbol{b}_i/\boldsymbol{b}_i'\boldsymbol{b}_i,\ i=1,\cdots,k-1,\ k=1,\cdots,q)$$

那么 $\{\boldsymbol{b}_i\}$ 是两两正交向量且 $\boldsymbol{b}_k$ 仅仅是前 k 个向量 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_k$ 的线性组合。再令

$$\boldsymbol{c}_i=\boldsymbol{b}_i/\|\boldsymbol{b}_i\|$$

于是 $\boldsymbol{c}_1,\cdots,\boldsymbol{c}_q$ 构成了 $\mathscr{L}(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q)$ 的一组标准正交基底。这种方法有时也称为 **Schmidt 正交化方法**。

此时如果记 $A=(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q)$，$C=(\boldsymbol{c}_1,\cdots,\boldsymbol{c}_q)$，由 Schmidt 正交化方法知道：存在矩阵$D$使得

$$C=AD \tag{1.4}$$

其中D是对角元素均大于 0 的上三角矩阵。

（四） 矩阵的秩和逆矩阵

对于给定的 $p\times q$ 阶矩阵 $A=(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q)$，称 $\mathscr{L}(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_q)$ 的维数为 A 的**秩**，记做 $\mathrm{rk}(A)$。从线性代数知识可知它有下列性质：

1) $\mathrm{rk}(A)=\mathrm{rk}(A')$。

2) $\mathrm{rk}(AB)\leqslant\min(\mathrm{rk}(A),\mathrm{rk}(B))$。

3) $\mathrm{rk}(A)=0$ 等价于 $A=0$。

4) 如果 A 是非零的 $p\times q$ 阶矩阵，则 $1\leqslant\mathrm{rk}(A)\leqslant\min(p,q)$。

5) $\max(\mathrm{rk}(A),\mathrm{rk}(B))\leqslant\mathrm{rk}(A\vdots B)\leqslant\min(p,\mathrm{rk}(A)+\mathrm{rk}(B))$。

6) $\mathrm{rk}\begin{pmatrix}A&0\\0&B\end{pmatrix}=\mathrm{rk}(A)+\mathrm{rk}(B)$。

设 A 是 p 阶方阵，定义它的**行列式**为

$$|A|=\sum_{\alpha}(-1)^{N(\alpha)}\prod_{j=1}^{p}a_{\alpha_j,j}$$

其中 $\sum\limits_{\alpha}$ 表示对数组 $\alpha=(\alpha_1,\cdots,\alpha_p)$ 的所有排列的加法，而 $N(\alpha)$ 表示排列 α 的逆序数，即将 $(\alpha_1,\cdots,\alpha_p)$ 变换成 $(1,2,\cdots,p)$ 需要的对换次数。例如 $\alpha=(3,4,1,2)$，那么 $N(\alpha)=4$。通常方阵 A 的行列式记作 $|A|$ 或者 $\det A$。

行列式具有下列性质：

1) $|I|=1$。

2) $|\Lambda|=\prod\limits_{i=1}^{p}\lambda_i$，其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_p)$。

3) $|A|=|A'|$。

4) $|A|=\prod\limits_{i=1}^{p}a_{ii}$，其中 A 是 p 阶上(下)三角矩阵，a_{ii} 是 A

的对角元素.

5) $|cA| = c^p|A|$. 其中 c 是常数, A 是 p 阶方阵.

6) 将方阵的任二行(列)予以交换,行列式改变符号.

7) 把矩阵的第 j 行(列)乘以一个常数加到第 i 行(列)上去,那么行列式的值不变. 于是当行列式中某行(列)是其他诸行(列)的线性组合时,行列式的值为0.

8) A,B 是同阶方阵,则 $|AB| = |A|\cdot|B|$.

9) $\left|\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\right| = |A|\cdot|B|$. 其中 A,B 均为方阵.

10) $|A\otimes B| = |A|^n\cdot|B|^p$, 其中 A 是 p 阶方阵. B 是 n 阶方阵.

证明　这里仅证明10)式

$$A\otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1p}B \\ \vdots & & \vdots \\ a_{p1}B & \cdots & a_{pp}B \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}I_n, & \cdots, & a_{1p}I_n \\ \vdots & & \vdots \\ a_{p1}I_n, & \cdots, & a_{pp}I_n \end{pmatrix}\begin{pmatrix} B & & & 0 \\ & B & & \\ & & \ddots & \\ 0 & & & B \end{pmatrix}$$

而

$$\begin{vmatrix} a_{11}I_n & \cdots & a_{1p}I_n \\ \vdots & & \vdots \\ a_{p1}I_n & \cdots & a_{pp}I_n \end{vmatrix} = |A|^n, \quad \left|\begin{pmatrix} B & & & 0 \\ & B & & \\ & & \ddots & \\ 0 & & & B \end{pmatrix}\right| = |B|^p.$$

所以10)式成立. ▮

令 $A_j^{(i)}$ 表示从方阵 A 中略去第 i 行和第 j 列后得到的子矩阵,称 $|A_j^{(i)}|$ 为 a_{ij} 的**余子式**,而称 $A(i,j) = (-1)^{i+j}|A_j^{(i)}|$ 为 a_{ij} 的**代数余子式**.

利用行列式性质6)可以得到

$$A(i,j)=\begin{vmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1p} \\ \vdots & & & & \\ a_{i-1,1} & & a_{i-1,j} & & a_{i-1,p} \\ 0\cdots0 & & 1 & & 0\cdots0 \\ a_{i+1,1} & \cdots & a_{i+1,j} & \cdots & a_{i+1,p} \\ \vdots & & & & \vdots \\ a_{p1} & \cdots & a_{pj} & \cdots & a_{pp} \end{vmatrix}$$

即 $A(i,j)$ 是把 A 中第 i 行改为 $a_{ij}=1$，其余 $a_{ik}=0(k\neq i, k=1,\cdots,p)$ 后得到的新方阵的行列式。于是有下列二式成立：

$$\sum_{j=1}^{p} a_{ij}A(k,j)=\begin{cases}|A| & k=i, \\ 0 & k\neq i.\end{cases}$$

$$\sum_{j=1}^{p} A(j,i)a_{jk}=\begin{cases}|A| & k=i, \\ 0 & k\neq i.\end{cases}$$

若记 $C=(A(i,j))$，上述结论则可表示为

定理 1.2 $C\cdot A'=A\cdot C'=|A|\cdot I$。

对于方阵 A，如果存在方阵 B 使得 $A\cdot B=B\cdot A=I$，则称 B 是 A 的**逆阵**，记为 $B=A^{-1}$，且它是唯一的。

如果方阵 A 的行列式 $|A|\neq0$，则称 A 是**非奇异方阵**。显然如果 A 有逆阵存在，则 $|A|\neq0$。反之若 A 是非奇异的，由定理 1.2 知道 $B=|A|^{-1}\cdot C'$ 正是 A 的逆阵。若记 $A^{-1}=(a^{ij})$，则有

$$a^{ij}=\frac{1}{|A|}A(j,i)$$

逆矩阵具有下列性质：

设 A,B 是 p 阶非奇异方阵，则

1) $(A^{-1})^{-1}=A$。

2) $(A')^{-1}=(A^{-1})'$。

3) $(A\cdot B)^{-1}=B^{-1}\cdot A^{-1}$。

4) $|A^{-1}|=|A|^{-1}$。

5) $|A| \neq 0 \Longleftrightarrow A$ 有逆矩阵 $\Longleftrightarrow \mathrm{rk}(A)=p$。

设 A,B 为非奇异方阵，则

6) $(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$.

7) 上(下)三角矩阵的逆阵仍为上(下)三角矩阵.

从性质 5) 知道非奇异方阵 A 的秩等于它的行(列)数，故我们也称 A 为**满秩矩阵**.

定理 1.3 1) 当 B 是非奇异方阵时，$\mathrm{rk}(AB)=\mathrm{rk}(A)$.

2) 令 B 是秩为 p 的 $p\times q$ 阶矩阵，则 $\mathrm{rk}(AB)=\mathrm{rk}(A)$.

证明 1) 记 $C=AB$，由于 B 非奇异，所以 $A=CB^{-1}$. 由秩的性质 2) 知道

$$\mathrm{rk}(C)\leqslant \mathrm{rk}(A)\leqslant \mathrm{rk}(C).$$

于是 $\mathrm{rk}(AB)=\mathrm{rk}(A)$.

3) 由 $\mathrm{rk}(B)=p$，知道 $q\geqslant p$，不妨设 $B=(B_1\vdots B_2)$ 其中 B_1 是 $p\times p$ 阶矩阵且 $\mathrm{rk}(B_1)=p$，于是由 1) 得到

$$\mathrm{rk}(A)=\mathrm{rk}(AB_1)\leqslant \mathrm{rk}(AB)\leqslant \mathrm{rk}(A)$$

从而 $\mathrm{rk}(AB)=\mathrm{rk}(A)$ ▌

推论 1.1 设 P、Q 非奇异，$PAQ=PBQ$，则 $A=B$.

定理 1.4 $\mathrm{rk}(A'A)=\mathrm{rk}(A)$.

证明 如果 A 为 $p\times q$ 阶矩阵且 A 的秩为 p，则由定理 1.3 知 $\mathrm{rk}(A'A)=\mathrm{rk}(A')=\mathrm{rk}(A)$. 一般情况，由 $Ax=0\to A'Ax=0\to x'A'Ax=0\to Ax=0$，从而 $\mathscr{L}(A)=\mathscr{L}(A'A)$，于是 $\mathrm{rk}(A'A)=\mathrm{rk}(A)$. ▌

设 A 是 p 阶方阵，其对角线元素的和称为 A 的**迹**. 记为

$$\mathrm{tr}(A)=\sum_{i=1}^{p} a_{ii}.$$

关于迹的运算有下列性质：

1) $\mathrm{tr}(\alpha A+\beta B)=\alpha\mathrm{tr}A+\beta\mathrm{tr}B$，其中 α、β 是常数. A、B 是同阶方阵.

2) 设 A 是 $p\times q$ 阶矩阵，B 是 $q\times p$ 阶矩阵，则

$$\mathrm{tr}(AB)=\mathrm{tr}(BA)=\sum_{i=1}^{p}\sum_{j=1}^{q}a_{ij}b_{ji}.$$

3) $\mathrm{tr}(A\otimes B)=\mathrm{tr}A\cdot\mathrm{tr}B$.

4) 设 A,B 都是 $p\times q$ 阶矩阵,则

$$\mathrm{tr}A'B=\mathrm{vec}A'\mathrm{vec}B=\sum_{i=1}^{p}\sum_{j=1}^{q}a_{ij}b_{ij}.$$

(五) 矩阵的分解

首先我们介绍初等阵和矩阵的**初等变换**的定义.

初等阵 $P_{ij},D_i(k),T_{ij}(k)$ 分别如下规定:

P_{ij} 是在单位阵 I 中改 $a_{ii}=a_{jj}=0$, $a_{ij}=a_{ji}=1$ 后得到.

$D_i(k)$ 是在单位阵 I 中改 $a_{ii}=k$ 后得到.

$T_{ij}(k)$ 是在单位阵 I 中改 $a_{ij}=k$ 后得到.

矩阵的初等变换是指对矩阵施行下列变换:

1) 对调矩阵两行(列)的位置.

2) 以不等于零的常数 k 乘以矩阵某一行(列).

3) 以非零常数 k 乘以某行(列)后加到另一行(列)的对应元素上去.

容易看出,上述变换正相当于分别用初等阵 $P_{ij},D_i(k),T_{ij}(k)$ 分别左(右)乘该矩阵.

由于初等阵是满秩的,且它们的逆阵是相应的初等阵,而初等阵的乘积仍是满秩阵. 于是有

定理 1.5 1) 设 A 是 $p\times q$ 阶矩阵, $\mathrm{rk}(A)=r$, 则存在 $p\times p$ 阶满秩矩阵 P 和 $q\times q$ 阶满秩矩阵 Q 使得

$$A=P\cdot\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}Q \tag{1.5}$$

2) 设 A 是 $p\times q$ 阶矩阵. $\mathrm{rk}(A)=r$, 则存在秩为 r 的矩阵 F 和 G, 其中 F 是 $p\times r$ 阶, G 是 $r\times q$ 阶,使得

$$A=F\cdot G$$

证明 1) 只须对 A 施行初等变换即可得证.

2）在(1.5)式中令 $P=(F, P_2)$，其中 F 是 $p\times r$ 阶矩阵，P_2 是 $p\times(p-r)$ 阶矩阵. 以及令 $Q=\begin{pmatrix}G\\Q_2\end{pmatrix}$，其中 G 是 $r\times q$ 阶矩阵，Q_2 是 $(q-r)\times q$ 阶矩阵，显然 $\mathrm{rk}(F)=\mathrm{rk}(G)=r$，立即可得

$$A=F\cdot G \quad \blacksquare$$

对于 p 阶方阵 U，如果它满足 $U\cdot U'=U'\cdot U=I_p$，则称 U 是 p 阶**正交矩阵**. 这时若记 $U=(u_1,\cdots,u_p)$，则有 $u_i'u_j=\delta_{ij}$. 这里

$$\delta_{ij}=\begin{cases}1 & i=j\\0 & i\neq j\end{cases}.$$

即 $\{u_1,\cdots,u_p\}$ 是 $\mathscr{L}(U)$ 的标准正交基底.

正交阵具有下列性质:

设 U_1,U_2 是同阶正交阵.

1）$U_1^{-1}=U_1'$.

2）$|U_1|=\pm 1$.

3）$U_1\cdot U_2, U_1^{-1}, U_1\otimes U_2$ 仍为正交阵.

定理 1.6 设 A 是 p 阶满秩矩阵，则它可唯一表示为 $A=U\cdot T$，其中 U 为 p 阶正交阵，T 是对角元素 $t_{ii}>0$ 的上三角矩阵.

证明 由于 A 满秩，A 的诸列向量线性无关. 由(1.4)式知道通过 Schmidt 正交化方法可把 A 表示为

$$A=U\cdot T$$

成立，其中 U 为 p 阶正交阵，T 为上三角矩阵且诸 $t_{ii}>0$. 设 $A=U_1\cdot T_1=U_2\cdot T_2$ 有不同的上述表示从而

$$U_2^{-1}U_1=T_2\cdot T_1^{-1}$$

即 $T_2T_1^{-1}$ 既是上三角阵又是正交阵，于是 $T_2T_1^{-1}=I$（见习题 1.13）. 即 $T_2=T_1$，因此 $U_1=U_2$，上述 A 的表示式是唯一的. $\blacksquare$

上述表示法也常称为 A 的**正交三角分解**.

定理 1.7 设 A 为 p 阶满秩阵，且诸 i 阶主子行列式均不为 0

$(i=1,\cdots,p)$，则存在上三角矩阵 U 和下三角矩阵 T，它们都是 p 阶满秩阵，使得

$$A=T\cdot U$$

证明　利用归纳法．当 $p=1$ 时显然成立．设 $p-1$ 时定理成立．记

$$A=\begin{pmatrix} A_{11} & v \\ t' & \omega \end{pmatrix}$$

其中 A_{11} 是 $p-1$ 阶满秩阵(由于 $|A_{11}|\neq 0$)，t, v 是 $p-1$ 维向量，ω 为实数．由归纳法 $A_{11}=T_1\cdot U_1$ (其中 T_1 为 $p-1$ 维上三角矩阵，U_1 为 $p-1$ 维上三角矩阵) 立即可得 $A=T\cdot U$．其中

$$T=\begin{pmatrix} T_1 & 0 \\ t'U_1^{-1} & 1 \end{pmatrix} \qquad U=\begin{pmatrix} U_1 & T_1^{-1}v \\ 0 & \omega-t'U_1^{-1}T_1^{-1}v \end{pmatrix}$$

仍分别为下三角矩阵和上三角矩阵且满秩．定理得证．▮

定理 1.6 和定理 1.7 得到的矩阵因子分解表示法是十分有用的．

§1.2　非奇异矩阵的分块逆矩阵及其一般矩阵的广义逆矩阵

本节将给出非奇异矩阵的分块矩阵的逆矩阵及其一般矩阵的广义逆矩阵的表达形式．

(一)　非奇异矩阵的分块矩阵求逆

设 A 是 p 阶满秩矩阵，它的逆矩阵为 A^{-1}，它们可表示为下列分块矩阵：

$$A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},\quad A^{-1}=\begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}$$

其中 A_{11}，A^{11} 均为 $r\times r$ 阶矩阵，A_{22} 和 A^{22} 均为 $s\times s$ 阶矩阵，$r+s=p$．

若 A_{11} 满秩，记 $A_{2\cdot 1}=A_{22}-A_{21}A_{11}^{-1}A_{12}$．

若 A_{22} 满秩，记 $A_{1\cdot2}=A_{11}-A_{12}A_{22}^{-1}A_{21}$。

定理 2.1 如果 A 和 A_{11} 非奇异，则

$$\begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}=\begin{pmatrix} A_{11}^{-1}+A_{11}^{-1}A_{12}A_{2\cdot1}^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{2\cdot1}^{-1} \\ -A_{2\cdot1}^{-1}A_{21}A_{11}^{-1} & A_{2\cdot1}^{-1} \end{pmatrix} \tag{2.1}$$

如果 A 和 A_{22} 非奇异，则

$$\begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}=\begin{pmatrix} A_{1\cdot2}^{-1} & -A_{1\cdot2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{1\cdot2}^{-1} & A_{22}^{-1}+A_{22}^{-1}A_{21}A_{1\cdot2}^{-1}A_{12}A_{22}^{-1} \end{pmatrix} \tag{2.2}$$

证明　仅证(2.1)式，(2.2)式可类似证明。

由于 A_{11}^{-1} 存在，容易验证下式成立：

$$\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}$$

$$=\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22}-A_{21}A_{11}^{-1}A_{12} \end{pmatrix}$$

由于 $|A|=|A_{11}|\cdot|A_{2\cdot1}|\neq0$，知 $|A_{2\cdot1}|\neq0$，于是 $A_{2\cdot1}^{-1}$ 存在，上式两边求逆并作适当变换得到：

$$\begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}=\begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}\begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{2\cdot1}^{-1} \end{pmatrix}\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}$$

展开后即得(2.1)式。 ▮

如果 A_{11}, A_{22} 都非奇异，则有

$$A_{1\cdot2}^{-1}=A_{11}^{-1}+A_{11}^{-1}A_{12}A_{2\cdot1}^{-1}A_{21}A_{11}^{-1} \tag{2.3}$$

$$A_{2\cdot1}^{-1}=A_{22}^{-1}+A_{22}^{-1}A_{21}A_{1\cdot2}^{-1}A_{12}A_{22}^{-1}. \tag{2.4}$$

另外还有

$$A^{11}=A_{1\cdot2}^{-1},\quad A^{22}=A_{2\cdot1}^{-1}$$

$$A^{12}=-A_{1\cdot2}^{-1}A_{12}A_{22}^{-1}=-A_{11}^{-1}A_{12}A_{2\cdot1}^{-1}$$

和

$$A^{21}=-A_{2\cdot1}^{-1}A_{21}A_{11}^{-1}=-A_{22}^{-1}A_{21}A_{1\cdot2}^{-1}$$

关于非奇异矩阵的行列式则有

推论 2.1 如果 A_{11}, A_{22} 非奇异，则

$$|A|=|A_{11}||A_{2\cdot1}|=|A_{22}||A_{1\cdot2}|.$$

推论 2.2 设 A 是 $p\times q$ 阶矩阵。B 是 $q\times p$ 阶矩阵，$\lambda\neq$

0，那么

$$|\lambda I_p - AB| = \lambda^{p-q}|\lambda I_q - BA| \tag{2.5}$$

证明　利用推论 2.1 可知

$$\begin{vmatrix} \lambda I_p & A \\ B & I_q \end{vmatrix} = |\lambda I_p|\left|I_q - \frac{1}{\lambda}BA\right| = \lambda^{p-q}|\lambda I_q - BA|$$

$$\begin{vmatrix} \lambda I_p & A \\ B & I_q \end{vmatrix} = |I_q||\lambda I_p - AB| = |\lambda I_p - AB|.$$ ▮

推论 2.3　设 A 是 $p\times q$ 阶矩阵，B 是 $q\times p$ 阶矩阵，那么

$$|I_p + AB| = |I_q + BA|$$

证明　只须在推论 2.2 中令 $\lambda=1$，A 用 $-A$ 代替即可. ▮

关于涉及矩阵的和的求逆下列定理是十分有用的.

定理 2.2　设 P 非奇异，$Q=P+UV$ 和 $I+VP^{-1}U$ 也非奇异，则

$$Q^{-1} = (P+UV)^{-1} = P^{-1} - P^{-1}U(I+VP^{-1}U)^{-1}\cdot V\cdot P^{-1} \tag{2.6}$$

证明　令 $A=\begin{pmatrix} P & -U \\ V & I \end{pmatrix}$，利用前面的记号，只须将

$A_{1\cdot2}=P+UV$，$A_{2\cdot1}=I+VP^{-1}U$，$A_{12}=-U$，$A_{21}=V$

代入 (2.3)式即可得证. ▮

例如，若 x 和 y 分别为 p 维向量，P 是 p 维方阵满足定理 2.2，则有

$$(P+x\cdot y')^{-1} = P^{-1} - (1+y'P^{-1}x)^{-1}\cdot P^{-1}x\cdot y^{-1}P^{-1} \tag{2.7}$$

*(二)　一般矩阵的广义逆矩阵

众所周知，逆矩阵在解线性方程组时十分有用，当 A 是 p 阶非奇异矩阵，线性方程组 $Ax=y$ 有唯一解. 且可求出解为 $x=A^{-1}y$. 这里 A^{-1} 是 A 的逆阵.

然而当 A 是奇异矩阵（即 $|A|=0$）或者当 A 不是方阵时却没有上述良好结果. 但我们仍然希望求得与 A^{-1} 类似的矩阵，使

得当 $Ax=y$ 相容时，它的解可以由 $x=Gy$ 表出，这样的矩阵 G 称为 A 的**广义逆**，记为 A^-，显然这时唯一性等性质将要丧失。本书对于广义逆仅要求掌握它的概念和理解它的一些表示方法。

定义 2.1 设 A 为 $p\times q$ 阶矩阵，A 的广义逆 A^- 是一个 $q\times p$ 阶矩阵，它使得对任一使方程 $Ax=y$ 相容的向量 y，$x=A^-y$ 是 $Ax=y$ 的解。

定理 2.3 A^- 存在的充分必要条件是 $AA^-A=A$ 成立。

证明 设 z 是任意 p 维向量，当 $y=Az$ 时，$Ax=y$ 相容，从而 A^- 存在保证了对任意向量 z 有

$$A(A^-A)z=A(A^-y)=y=Az$$

于是 $AA^-A=A$。

反之，若 $Ax=y$ 相容，意味着满足 $Aw=y$ 的向量 w 存在，由于 $AA^-A=A$，所以有

$$AA^-(Aw)=Aw$$

即 $A(A^-y)=y$，这说明 A^-y 是 $Ax=y$ 的解。 ∎

定义 2.2 对任何矩阵 A，A^- 总存在。且若 $\mathrm{rk}(A)=r$ 且 A 可表示为

$$A=P\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}Q \quad（其中\ P,Q\ 为满秩矩阵）$$

则 A^- 的一般形式是

$$A^-=Q^{-1}\begin{pmatrix}I_r & C\\ D & E\end{pmatrix}P^{-1} \tag{2.8}$$

其中 C,D,E 是任意的相应的子矩阵。

证明 利用 $AA^-A=A$ 知道

$$P\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}QA^-P\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}Q=P\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}Q$$

由推论 1.1 有

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} QA^- P \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

记

$$QA^- P = \begin{pmatrix} B & C \\ D & E \end{pmatrix}$$

代入上式，得到 $B = I_r$，C, D, E 任意. 从而(2.8)式成立. ▮

由 A^- 的一般表达式(2.8)，我们可以得到下列结论：

1) A 的广义逆一般有无穷多个，而 A^- 唯一的充分必要条件是 A^{-1} 存在，且 $A^- = A^{-1}$.

2) $\mathrm{rk}(A^-) \geqslant \mathrm{rk}(A)$,

结论 2) 可由 $\mathrm{rk}(A^-) \geqslant \mathrm{rk}(AA^-A) = \mathrm{rk}(A)$ 得到.

下面介绍广义逆与幂等矩阵之间的联系.

如果矩阵 A 满足 $A^2 = A$，那么称 A 为**幂等矩阵**(或幂等阵).

3) AA^- 和 A^-A 是幂等矩阵，且 $\mathrm{rk}(AA^-) = \mathrm{rk}(A^-A) = \mathrm{rk}(A)$.

证明　由 $AA^-A = A$ 立即可知 AA^- 和 A^-A 是幂等矩阵. 结论的第二部分可由

$$\mathrm{rk}(A) = \mathrm{rk}(AA^-A) \leqslant \mathrm{rk}(AA^-) \leqslant \mathrm{rk}(A)$$

得到. ▮

定理 2.5　1) $(A'A)^-A'$ 是 A 的广义逆，$A'(AA')^-$ 也是 A 的广义逆.

2) $A(A'A)^-A'A = A$，$AA'(AA')^-A = A$.

3) $P = A(A'A)^-A'$ 和 $Q = A'(AA')^-A$ 是幂等对称矩阵，且 $\mathrm{rk}(P) = \mathrm{rk}(Q) = \mathrm{rk}(A)$. 且它们是唯一的.

证明　类似于定理 1.4 知道矩阵方程

$$AX = 0 \rightarrow A'AX = 0 \rightarrow X'A'AX = 0 \rightarrow AX = 0$$

所以有

$$A'AX = A'AY \longleftrightarrow AX = AY.$$

因此

$$A'A(A'A)^-A'A = A'A \longleftrightarrow A(A'A)^-A'A = A$$

从而 1),2) 的第一式得证。类似可证第二式。

由广义逆性质知 $P = A(A'A)^-A'$ 是幂等阵，且 $\mathrm{rk}(P) = \mathrm{rk}(A)$。记 $P_i = A(A'A)_i^-A' (i = 1,2)$，其中 $(A'A)_1^-$ 和 $(A'A)_2^-$ 是 $A'A$ 的任意二个不同的广义逆，那么

$$\begin{aligned} P_1 &= A(A'A)_1^-A' = [A(A'A)_2^-A'A](A'A)_1^-A' \\ &= A(A'A)_2^-[A'A(A'A)_1^-A'] = A(A'A)_2^-A' \\ &= P_2 \end{aligned}$$

即 $A(A'A)^-A'$ 关于 $A'A$ 的任一广义逆是唯一的。

将等式 $A'A(A'A)^-A'A = A'A$ 转置后知道 $[(A'A)^-]'$ 也是 $A'A$ 的广义逆。所以

$$P' = [A(A'A)^-A']' = A[(A'A)^-]'A' = P.$$

即 P 是对称幂等阵。关于 Q 可类似证明。 ▌

关于奇异方阵的广义逆矩阵的分块表示形式与非奇异时求得的表达式类似，只是多出一些任意项。读者感兴趣可参考[9]。

由于广义逆的形式不唯一，给运用带来了一些不便。我们希望从众多的广义逆中找出一个具有"良好"性质的特殊广义逆。我们称满足下列性质的广义逆为 A^+：

1) $AA^+A = A$.　　　　2) $A^+AA^+ = A^+$.

3) $(AA^+)' = AA^+$.　　　　4) $(A^+A)' = A^+A$.

称 A^+ 为 A 的**加号逆**，又称 A 的 **Moore-Penrose 逆**。

对于任一个秩为 r 的 $p \times q$ 阶矩阵 A，按定理 1.5 知道存在秩为 r 的 $p \times r$ 阶矩阵 F 和 $r \times q$ 阶矩阵 G，使得 $A = FG$ 成立。这时 $F'F$ 及 GG' 的逆阵存在，那么

$$B_0 = G'(GG')^{-1}(F'F)^{-1}F'$$

就是满足上述条件的唯一的 A^+。

B_0 满足 A^+ 的四条性质是容易证明的(见习题 1.4)，下面我们证明它是唯一的。设 C 也满足上述四条性质，则有

$$B_0A = B_0A \cdot CA = A'B_0' \cdot A'C' = (AB_0A)' \cdot C' = A'C' = CA$$

同理可证 $AB_0=AC$。于是

$$B_0=B_0AB_0=CAB_0=CAC=C.$$

即 A^+ 是唯一的。 ▌

下面给出 A^+ 的一些基本性质，请读者自行验证。

1) $(A^+)^+=A$, $(A')^+=(A^+)'$.

2) $A^+=(A'A)^+A'=A'(AA')^+$.

3) 若 A 是对称幂等阵即 $A^2=A$, $A'=A$, 则 $A^+=A$.

4) A 是满秩方阵，则 $A^+=A^{-1}$.

5) 设 P_1,P_2 是正交阵，则 $(P_1AP_2)^+=P_2'A^+P_1'$.

6) 设 $A=\mathrm{diag}(\lambda_1,\cdots,\lambda_p)$ 则 $A^+=\mathrm{diag}(\lambda_1^+,\cdots,\lambda_p^+)$，其中

$$\lambda_i^+=\begin{cases}\lambda_i^{-1}, & \lambda_i\neq 0\\ 0, & \lambda_i=0.\end{cases}$$

例 求矩阵

$$A=\begin{pmatrix}1 & 0 & 1\\ -1 & 1 & 0\\ 0 & 1 & 1\end{pmatrix}$$

的加号逆 A^+.

解 按定理 1.5 $A=F\cdot G$ 其中

$$F=\begin{pmatrix}1 & 0\\ 0 & 1\\ 1 & 1\end{pmatrix},\quad G=\begin{pmatrix}1 & 0 & 1\\ -1 & 1 & 0\end{pmatrix}$$

于是

$$(F'F)^{-1}=\frac{1}{3}\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix}\quad (GG')^{-1}=\frac{1}{3}\begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix}$$

所以

$$A^+=G'(GG')^{-1}(F'F)^{-1}F'$$

$$=\frac{1}{3}\begin{pmatrix}1 & -1 & 0\\ 0 & 1 & 1\\ 1 & 0 & 1\end{pmatrix}$$

注 在上例中由于 A 的秩为 2，它的第一，二行线性无关，故可直接取它为上述 G，而 F 可由 $A=FG$ 直接求得。一般地 F，G 可利用定理 1.5 求出。

§1.3 若干特殊矩阵及其性质

本节中我们将讨论对称阵，幂等阵，对称幂等阵(投影阵)以及正定矩阵的有关性质并给出二次型极值的一些性质。

（一） 对称幂等阵及其性质

首先我们复习一些有关特征根和特征向量的知识。

设 A 是 p 阶方阵，若 $\boldsymbol{x}$ 是 p 维非零向量，满足 $A\boldsymbol{x}=\lambda\boldsymbol{x}$，其中 λ 是一个实数。则 λ 称为 A 的**特征根**，而 $\boldsymbol{x}$ 称为 A 相应于 λ 的**特征向量**。由于 $(A-\lambda I_p)\boldsymbol{x}=0$，所以 $\mathrm{rk}(A-\lambda I_p)\leqslant p-1$，即 λ 是**特征方程** $|A-\lambda I_p|=0$ 的根。反之，若 $|A-\lambda I_p|=0$ 有根 λ，则保证存在非零向量 $\boldsymbol{x}$ 使得 $A\boldsymbol{x}=\lambda\boldsymbol{x}$，这说明任一方阵总有特征根和特征向量。

由线性代数我们知道

定理 3.1 若 A 是 $p\times p$ 阶实对称阵，则

1）A 的特征根全为实数。

2）相应于不同的特征根的 A 的特征向量相互正交。

3）存在一个正交矩阵 $P=(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_p)$，使得

$$P'AP=\mathrm{diag}(\lambda_1,\cdots,\lambda_p)$$

其中 λ_i 是 A 的特征根 $(i=1,\cdots,p)$。

由定理 3.1 知道 A 可写成下列形式

$$A=\sum_{i=1}^{p}\lambda_i\boldsymbol{e}_i\boldsymbol{e}_i'$$

这种表达式称为 A 的**谱分解**。

为方便起见，以后我们总以 $\lambda_i(A)=\mathrm{ch}_i(A)$ 表示 A 的第 i 大的特征根 $(i=1,\cdots,p)$。

定理 3.2 设 A 是 $p\times p$ 阶矩阵且 $\mathrm{rk}(A)=r$，则

1) A 为幂等阵的充分必要条件是 $\mathrm{rk}(A)+\mathrm{rk}(I_p-A)=p$.

2) 幂等阵的特征根只能是 0 和 1.

3) 如果 $A=A^2$, 则 $\mathrm{tr}(A)=\mathrm{rk}(A)$.

4) 如果 A 为对称幂等阵,则存在 $p\times r$ 阶矩阵 U, 使得

$$A=U\cdot U',\ 且\ U'U=I_r.$$

证明　由于

$$\begin{pmatrix}I_p & -I_p\\ 0 & I_p\end{pmatrix}\begin{pmatrix}I_p & I_p-A\\ A & 0\end{pmatrix}\begin{pmatrix}I_p & 0\\ -I_p & I_p\end{pmatrix}=\begin{pmatrix}0 & I_p-A\\ A & 0\end{pmatrix},$$

$$\begin{pmatrix}I_p & 0\\ -A & I_p\end{pmatrix}\begin{pmatrix}I_p & I_p-A\\ A & 0\end{pmatrix}\begin{pmatrix}I_p & -(I_p-A)\\ 0 & I_p\end{pmatrix}$$

$$=\begin{pmatrix}I_p & 0\\ 0 & -(A-A^2)\end{pmatrix}$$

所以 $\mathrm{rk}(A-A^2)+p=\mathrm{rk}(A)+\mathrm{rk}(I_p-A)$, 由此立即推出 1).

如果 λ 是幂等阵 A 的特征根,则存在非零向量 $\boldsymbol{x}$, 使得

$$\lambda\boldsymbol{x}=A\boldsymbol{x}=A^2\boldsymbol{x}=\lambda(A\boldsymbol{x})=\lambda^2\boldsymbol{x}$$

因而 $\lambda^2=\lambda$, λ 非 0 即 1, 这就证明了 2).

由于 $\mathrm{rk}(A)=r$, 由定理 1.5 知存在满秩矩阵 P 和 Q, 使得

$$A=P\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}Q$$

由 $A^2=A$ 以及推论 1.1 可知

$$\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}QP\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}=\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}$$

所以

$$\mathrm{tr}(A)=\mathrm{tr}\left[P\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}Q\right]$$

$$=\mathrm{tr}\left[\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}QP\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}\right]$$

$$= \mathrm{tr}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = r = \mathrm{rk}(A).$$

于是 3) 得证.

如果 A 是 p 阶对称幂等阵，由定理 3.1 的 3) 知道存在正交矩阵 $P = (U, V)$，(其中 U 是 $p \times r$ 阶，V 是 $p \times (p - r)$ 阶)使得

$$A = P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} P' = U \cdot U'$$

且 $U'U = I_r$ 成立. ▌

定理 3.3 若 $p \times p$ 阶矩阵 A 是对称幂等阵且 $\mathrm{rk}(A) = r$，那么一定存在秩为 r 的 $p \times r$ 阶矩阵 P_1，使得

$$A = P_1(P_1'P_1)^{-1}P_1'$$

证明 由 $\mathrm{rk}(A) = r$，知存在秩为 r 的 $p \times r$ 阶矩阵 P_1 和 Q_1 使得 $A = P_1Q_1'$. 由 $A^2 = A$ 推得

$$P_1(Q_1'P_1 - I_r)Q_1' = 0$$

由推论 1.1 知，$Q_1'P_1 = I_r$ 成立. 又由 A 对称，所以 $P_1Q_1' = Q_1P_1'$，从而

$$P_1 = P_1 \cdot Q_1'P_1 = Q_1 \cdot P_1'P_1$$

由于 $P_1'P_1$ 是 r 阶满秩矩阵，$(P_1'P_1)^{-1}$ 存在，所以

$$Q_1 = P_1(P_1'P_1)^{-1}.$$

于是 $A = P_1(P_1'P_1)^{-1}P_1'$. ▌

上述讨论的对称幂等阵 A 很有用，它有许多优良性质. 如它也是非负定矩阵，$A^+ = A$，且存在正交阵 P 使得

$$P'AP = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}, \quad r = \mathrm{rk}(A)$$

等等，下面介绍对称幂等阵在几何上的意义.

设 $\mathscr{L}$ 是线性空间 R^n 上的一个线性子空间，则一定存在另一子空间 $\mathscr{L}^\perp$，使得对每一个向量 $x \in R^n$，x 可唯一分解成 $x = x_1 + x_2$，其中 $x_1 \in \mathscr{L}$，$x_2 \in \mathscr{L}^\perp$，且对于任意的 $x_1 \in \mathscr{L}$ 和

$x_2 \in \mathscr{L}^{\perp}$ 有 $x_1'x_2 = 0$，我们称 $\mathscr{L}^{\perp}$ 是 $\mathscr{L}$ 关于 R^n 的**正交补空间**，而 x_1 称为 x 在 $\mathscr{L}$ 上的**投影**。显然当 x_1 和 y_1 是 x, y 在 $\mathscr{L}$ 上的投影，那么 $ax_1 + by_1$ 正是 $ax + by$ 的投影，即投影算子是一个齐次线性算子。而在有限维空间中，投影可由一个矩阵来实现，即 $x_1 = Px$。其中 P 取决于 $\mathscr{L}$，称 P 为 $\mathscr{L}$ 上的**投影矩阵**，简称投影阵。

定理 3.4 矩阵 P 是投影矩阵的充分必要条件是 P 为对称幂等阵。

证明 若 P 是 $\mathscr{L}$ 上的投影矩阵，则对于一切 $x \in R^n$ 有

$$P^2x = Px_1 = x_1 = Px$$

即 $P^2 = P$。又由 $(x - x_1)'x_1 = 0$ 得到

$$x'(I - P)'Px = x'(P - P'P)x = 0$$

对一切 $x \in R^n$ 成立，所以 $P = P'P$ 以及 $P = P'$ 成立，从而知道投影阵 P 是对称幂等阵。

反之，若 P 是对称幂等阵，由定理 3.3 知存在 P_1，使得 $P = P_1(P_1'P_1)^{-1}P_1$. 记 $\mathscr{L} = \mathscr{L}(P_1)$，于是

$$x_1 = Px = P_1\xi \in \mathscr{L}(P_1) \quad 其中 \ \xi = (P_1'P_1)^{-1}P_1'x.$$

而 $x_2 = (I - P)x$ 满足

$$x_2'x_1 = x'(I - P)'Px = 0$$

即 $x_2 \in \mathscr{L}^{\perp}(P_1)$，$x_1$ 是 x 在 $\mathscr{L}(P_1)$ 上的投影阵，即 P 是 $\mathscr{L}(P_1)$ 上的投影阵。 ▮

以后为方便起见我们也称对称幂等阵为投影阵。

定理 3.5 设 A 是 $p \times q$ 阶矩阵，则 $P = A(A'A)^{-}A'$ 是 $\mathscr{L}(A)$ 上的投影阵。

证明 由定理 2.5 知 P 是对称幂等阵。又对于任一向量 $x \in R^n$ 有 $Px = A\xi \in \mathscr{L}(A)$（其中 $\xi = (A'A)^{-}A'x$）。

如同定理 3.4 证明可知 P 是 $\mathscr{L}(A)$ 上的投影阵。 ▮

注意，对于同一子空间 $\mathscr{L}(A)$ 上的投影阵可以有几种不同的表示方法，但它们实质上是同一矩阵。以后我们常用 $P_A =$

$A(A'A)^{-}A'$ 表示 $\mathscr{L}(A)$上的投影阵,此时 $I-P_A$ 表示 $\mathscr{L}^{\perp}(A)$ 上的投影阵。

定理 3.6 设 $\mathscr{L}$ 是 R^n 中子空间, 向量 $a\in R^n$, 则在 $\mathscr{L}$ 中存在向量 b 使得

$$\|b-a\|=\inf_{x\in\mathscr{L}}\|x-a\|$$

证明 令 P 是 $\mathscr{L}$ 上的投影阵, 则 $a=Pa+(I-P)a$, 其中 $Pa\in\mathscr{L}$, $(I-P)a\in\mathscr{L}^{\perp}$. 由于对任一 $x\in\mathscr{L}$ 有 $x'(I-P)a=0$。于是

$$(x-a)'(x-a)=(x-Pa)'(x-Pa)+\|(I-P)a\|^2$$

显然当取 $b=Pa$ 时定理成立。 ▌

定理 3.6 指出对于任一向量 $a\in R^n$, 在 $\mathscr{L}$ 中以它的投影 Pa 在距离意义下与 a 最接近,它的几何意义十分直观.

下面定理指出投影阵可以进一步分解成两个投影阵的和, 它对于今后的讨论是有用的,尽管它的形式比较复杂.

定理 3.7 设秩为 p 的 $n\times p$ 阶矩阵 $X=(X_1,X_2)$, 其中 X_1 是 $n\times r$ 阶, X_2 是 $n\times(p-r)$ 阶,那么有

$$\begin{aligned}X(X'X)^{-1}X'=&X_2(X_2'X_2)^{-1}X_2'+(X_1-X_2(X_2'X_2)^{-1}X_2'X_1)\\&\cdot[X_1'X_1-X_1'X_2(X_2'X_2)^{-1}X_2'X_1]^{-1}(X_1-X_2(X_2'X_2)^{-1}X_2'X_1)'\end{aligned}\tag{3.1}$$

若记 $Y=(I-P_{X_2})X_1$, 上式可简记为

$$P_X=P_{X_2}+P_Y\tag{3.2}$$

即 $\mathscr{L}(X)$ 上的投影阵可分解为 $\mathscr{L}(X_2)$ 和 $\mathscr{L}(Y)$ 上投影阵之和。

证明 由于 $\mathrm{rk}(X)=p$, 故 $(X'X)^{-1}$, $(X_2'X_2)^{-1}$ 存在. 记

$$V=XX'=\begin{pmatrix}V_{11}&V_{12}\\V_{21}&V_{22}\end{pmatrix}$$

那么 $V_{11}=X_1'X_1$, $V_{22}=X_2'X_2$, $V_{21}=X_2'X_1$, $V_{12}=X_1'X_2$ 以及

$$V_{1\cdot2}=X_1'[I-X_2(X_2'X_2)^{-1}X_2']X_1=X_1'(I-P_{X_2})X_1$$

利用定理 2.1 的证明知道

$$V^{-1}=\begin{pmatrix} I & 0 \\ -V_{22}^{-1}V_{21} & I \end{pmatrix}\begin{pmatrix} V_{1\cdot 2}^{-1} & 0 \\ 0 & V_{22}^{-1} \end{pmatrix}\begin{pmatrix} I & -V_{12}V_{22}^{-1} \\ 0 & I \end{pmatrix}$$

于是通过运算求得

$$\begin{aligned} P_X &= (X_1, X_2)V^{-1}\begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \\ &= (X_1 - X_2V_{22}^{-1}V_{21})\cdot V_{1\cdot 2}^{-1}(X_1 - X_2V_{22}^{-1}V_{21})' + X_2V_{22}^{-1}X_2'. \end{aligned}$$

而 $X_1 - X_2V_{22}^{-1}V_{21}$ 正是 $Y=(I-P_{X_2})X_1$ 它满足 $X_2'Y=0$ 且 $V_{1\cdot 2}=Y\cdot Y'$ 因此(3.1)和(3.2)成立。▌

注意到定理 3.7 中 $\mathscr{L}(Y)$ 正是 $\mathscr{L}(X_1)$ 关于 $\mathscr{L}(X_2)$ 的正交补空间。若 $p=2$ 我们可以如下直观说明：此时 $\mathscr{L}(X)$ 表示由向量 x_1 和 x_2 张成的子空间，那么任一向量 $x\in R^n$ 在 $\mathscr{L}(X)$ 上的投影可以分解成两个投影之和，一个是 x 在 x_2 方向上的投影，另一个是在 $\mathscr{L}(X)$ 中与 x_2 垂直方向上的投影，即在 $y=(I-P_{x_2})x_1$ 方向上的投影。因此 $P_Xx=P_{x_2}x+P_yx$。换句话说：$P_X=P_{x_2}+P_y$。

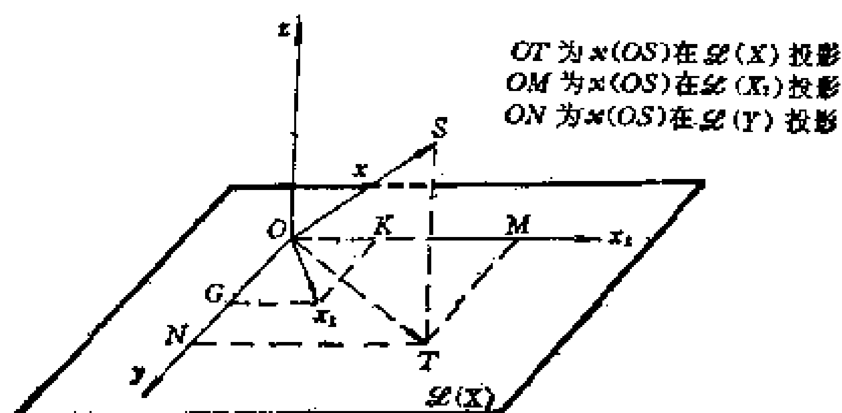


图 3.1

(二) 非负定矩阵及其性质

令 A 是 n 阶对称阵，x 是一个 n 维向量，那么 $x'Ax$ 是 x 诸分量的二次齐次多项式，简称为 x 的**二次型**.

如果对于所有非零向量 x 总有 $x'Ax>0$ ($\geqslant 0$)，那么称 A 是**正定矩阵**（**非负定矩阵**或者**半正定矩阵**）。简记为 $A>0$

$(A \geqslant 0)$.

定理 3.8 如果A是正定(半正定)矩阵时,A的所有特征根都为正(非负),诸主子矩阵 A_{ii} 也是正定(半正定)阵. 且 $|A|$ 以及A的诸主子行列式全为正(非负).

证明 由A对称及定理 3.1 知存在正定矩阵P使得

$$A = P'\mathrm{diag}(\lambda_1, \cdots, \lambda_p) \cdot P$$

令 $\boldsymbol{y} = P\boldsymbol{x} = (y_1, \cdots, y_p)'$, (其中 $\boldsymbol{x} = (x_1, \cdots, x_p)'$), 则

$$\boldsymbol{x}'A\boldsymbol{x} = (P\boldsymbol{x})'\mathrm{diag}(\lambda_1, \cdots, \lambda_p)P\boldsymbol{x} = \sum_{i=1}^{p} \lambda_i y_i^2 > 0(\geqslant 0)$$

对一切非零向量$\boldsymbol{y}$成立. 特别地,取 $y = (0, \cdots, 0, 1, 0, \cdots, 0)'$, 其中 1 在第 i 位置上,即得 $\lambda_i > 0(\lambda_i \geqslant 0)$. 此外

$$|A| = |P'\mathrm{diag}(\lambda_1, \cdots, \lambda_p)P| = \prod_{i=1}^{p} \lambda_i > 0(\geqslant 0).$$

若记由A的前 r 行,前 r 列组成的子矩阵为 A_{11}, 如果取 $\boldsymbol{y} = (\boldsymbol{z}', \boldsymbol{0}')'$, 其中 $\boldsymbol{z} = (z_1, \cdots, z_r)'$ 是 r 维向量, $\boldsymbol{0}$ 为诸元素全为 0 的 $p-r$ 维向量,那么 $\boldsymbol{z}'A_{11}\boldsymbol{z} = \boldsymbol{y}'A\boldsymbol{y}$ 对于一切非零向量 $\boldsymbol{z}$ 恒大于 0(恒大于等于 0),从而 A_{11} 正定(半正定),且 $|A_{11}| > 0\ (\geqslant 0)$ 成立. 对其他主子矩阵可类似证明. ▌

定理 3.9 一个矩阵A是半正定的充分必要条件是它可以表达成 $A = X \cdot X'$. 若A是正定的,则相应的X是非奇异方阵.

证明 充分性: 若 $A = XX'$, 则对于任一 p 维非零向量 $\boldsymbol{x}$,

$$\boldsymbol{x}'A\boldsymbol{x} = \boldsymbol{x}'X \cdot X'\boldsymbol{x} = \|\boldsymbol{x}'X\|^2 \geqslant 0$$

从而A半正定,若X非奇异时, 则当 $\boldsymbol{x} \neq 0$ 时 $X'\boldsymbol{x} \neq 0$, 上式恒大于 0, A是正定的.

必要性: 若A半正定,A的特征根为 $\lambda_1, \cdots, \lambda_r, 0, \cdots, 0$, 其中 $\lambda_i > 0(i = 1, \cdots, r)$. 于是存在 $p \times p$ 正交阵 $P = (P_1, P_2)$. 其中 P_1 是 $p \times r$ 阶,P_2 是 $p \times (p-r)$ 阶. 使得

$$A = P\mathrm{diag}(\lambda_1, \cdots, \lambda_r, 0, \cdots, 0)P' = P_1\mathrm{diag}(\lambda_1, \cdots, \lambda_r)P_1'$$

记 $X = P_1 \cdot \mathrm{diag}(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_r})$，则 $A = XX'$ 成立。若 A 正定，则 $X = P \cdot \mathrm{diag}(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_p})$ 是非奇异的。▌

推论 3.1 设 A 是正定矩阵，则 A^{-1} 也是正定矩阵。

证明 按照定理 3.8 的证明，可知

$$A^{-1} = P'\mathrm{diag}(\lambda_1^{-1}, \cdots, \lambda_p^{-1})P.$$

其中 P 正交矩阵，诸 $\lambda_i > 0$，从而 A^{-1} 显然正定。▌

定理 3.10 设 $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ 是 p 阶对称阵，A_{11} 是 r 阶非奇异对称阵 $(r \leqslant p)$，则 A 是正定的充分必要条件是 A_{11} 及 $A_{2\cdot1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$ 正定。

证明 令 $x = (x_1', x_2')'$，其中 x_1 是 r 维向量，x_2 是 $p-r$ 维向量。由于

$$A = \begin{pmatrix} I & 0 \\ A_{21}A_{11}^{-1} & I \end{pmatrix}\begin{pmatrix} A_{11} & 0 \\ 0 & A_{2\cdot1} \end{pmatrix}\begin{pmatrix} I & A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}.$$

$$x'Ax = (x_1 + A_{11}^{-1}A_{12}x_2)'A_{11}(x_1 + A_{11}^{-1}A_{12}x_2) + x_2'A_{2\cdot1}x_2.$$

如果 A_{11}、$A_{2\cdot1}$ 正定，则 A 显然正定。反之，若 A 正定，由定理 3.8 的证明知 A_{11} 也正定。对于任给的非零向量 x_2，只须令 $x_1 = -A_{11}^{-1}A_{12}x_2$，可知 $x_2'A_{2\cdot1}x_2 > 0$。故 $A_{2\cdot1}$ 正定。▌

在定理 3.8 的证明中我们有：正定阵 A 可表示为

$$A = P'\mathrm{diag}(\lambda_1, \cdots, \lambda_p)P.$$

若取 $B = P'\mathrm{diag}(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_p})P$，则 B 是正定阵，且 $A = B^2$。以后将称上述 B 是 A 的平方根矩阵，记为 $B = A^{\frac{1}{2}}$。而 B 的逆阵记为 $A^{-\frac{1}{2}}$。

上面的诸定理讨论了单个正定矩阵的性质及其它的充分必要条件，这些性质经常用到。下面将考虑与两个矩阵有关的特征根和特征向量的概念及性质。

设 A 是对称阵，$B > 0$。称 $|A - \lambda B| = 0$ 的根 λ 为 **A 相对于 B 的特征根**。相应于该 λ 而满足 $(A - \lambda B)x = 0$ 的向量

x 称为**对应于 λ 的 A 相对于 B 的特征向量.**

由于

$$|A-\lambda B|=0 \longleftrightarrow |B^{-\frac{1}{2}}AB^{-\frac{1}{2}}-\lambda I|=0$$

所以求 A 相对于 B 的特征根就是求 $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$ 的特征根.

设 x 是 $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$ 相应于特征根 λ 的特征向量,即

$$B^{-\frac{1}{2}}AB^{-\frac{1}{2}}x=\lambda x$$

那么

$$A(B^{-\frac{1}{2}}x)=\lambda B(B^{-\frac{1}{2}}x)$$

即 $B^{-\frac{1}{2}}x$ 正是相应于 λ 的 A 相对于 B 的特征向量.

定理 3.11 设 A 是对称阵, B 是同阶正定阵,则一定存在一个非奇异矩阵 R, 使得

$$B=RR' \quad A=RD_\lambda R'$$

其中 D_λ 是由 A 相对于 B 的诸特征根 λ_i 构成的对角阵. 当 A 半正定时,诸 λ_i 非负.

证明 由于 $B>0$, $B^{\frac{1}{2}}$ 和 $B^{-\frac{1}{2}}$ 存在且正定,而 $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$ 是对称阵,从而存在正交阵 P 使得

$$B^{-\frac{1}{2}}AB^{-\frac{1}{2}}=PD_\lambda P'$$

其中 D_λ 是 $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$ 的特征根也即 A 相对于 B 的特征根构成的对角矩阵. 令 $R=B^{\frac{1}{2}}P$, 则 R 显然非奇异且

$$B=RR' \quad A=RD_\lambda R'$$

成立. 而当 A 半正定时, $B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$ 也是半正定, 从而诸 λ_i 非负. ▮

定理 3.12 设 A 是 p 阶正定阵, 一定存在一个对角元素为正的下三角矩阵 T, 使得 $A=T\cdot T'$.

证明 由定理 3.9 知 $A=X\cdot X'$, 其中 X 是非奇异矩阵, 又由定理 1.6 知 X' 可表示为 $X'=U\cdot R$, 其中 U 为正交矩阵, R 是对角元素为正的上三角矩阵. 于是 $A=R'R$, 令 $T=R'$, 定理得证. ▮

在本段的最后部分我们集中讨论与二次型极值有关的不等

式，它们都有广泛的运用。

引理 3.1 (Cauchy-Schwarz 不等式)

设 $\boldsymbol{b}$ 和 $\boldsymbol{d}$ 是两个 p 维向量，则

$$(\boldsymbol{b}'\boldsymbol{d})^2 \leqslant (\boldsymbol{b}'\boldsymbol{b})(\boldsymbol{d}'\boldsymbol{d}) \tag{3.3}$$

上述等号当且仅当 $\boldsymbol{b} = c\boldsymbol{d}$(或 $\boldsymbol{d} = c\boldsymbol{b}$)时成立，这里 c 为常数。

证明　当 $\boldsymbol{b} = \boldsymbol{0}$ 或 $\boldsymbol{d} = \boldsymbol{0}$ 显然成立。不妨设 $\boldsymbol{b} \neq \boldsymbol{0}, \boldsymbol{d} \neq \boldsymbol{0}$，考虑向量 $\boldsymbol{b} - x\boldsymbol{d}$，这里 x 是一个变数，于是

$$\begin{aligned} 0 &\leqslant (\boldsymbol{b} - x\boldsymbol{d})'(\boldsymbol{b} - x\boldsymbol{d}) \\ &= (\boldsymbol{d}'\boldsymbol{d})x^2 - (2\boldsymbol{b}'\boldsymbol{d})x + \boldsymbol{b}'\boldsymbol{b} \end{aligned} \tag{3.4}$$

从 x 的二次函数性质知道此时

$$(\boldsymbol{b}'\boldsymbol{d})^2 - (\boldsymbol{b}'\boldsymbol{b})(\boldsymbol{d}'\boldsymbol{d}) \leqslant 0.$$

即(3.3) 式成立。

若 $\boldsymbol{b} = c\boldsymbol{d}$，立即可知(3.3)式等号成立。反之若 (3.3) 式等号成立，则当取 $x = \boldsymbol{b}'\boldsymbol{d}/\boldsymbol{d}'\boldsymbol{d}$ 时(3.4)式为 0，从而存在常数 $c = \boldsymbol{b}'\boldsymbol{d}/\boldsymbol{d}'\boldsymbol{d}$ 使得 $\boldsymbol{b} = c\boldsymbol{d}$。 ▮

引理 3.2 (推广的 Cauchy-Schwarz 不等式)

设 $\boldsymbol{b}, \boldsymbol{d}$ 是两个 p 维向量，B 是 p 阶正定矩阵，那么有

$$(\boldsymbol{b}'\boldsymbol{d})^2 \leqslant (\boldsymbol{b}'B\boldsymbol{b})(\boldsymbol{d}'B^{-1}\boldsymbol{d}) \tag{3.5}$$

且等号当且仅当 $\boldsymbol{b} = cB^{-1}\boldsymbol{d}$ (或 $\boldsymbol{d} = cB\boldsymbol{b}$) 时成立，这里 c 常数。

证明　由于 B 正定，记 $\tilde{\boldsymbol{b}} = B^{\frac{1}{2}}\boldsymbol{b}$，$\tilde{\boldsymbol{d}} = B^{-\frac{1}{2}}\boldsymbol{d}$，此时 $\boldsymbol{b}$ 与 $\tilde{\boldsymbol{b}}$，$\boldsymbol{d}$ 与 $\tilde{\boldsymbol{d}}$ 同时为零向量或者非零向量，利用引理 3.1 于向量 $\tilde{\boldsymbol{b}}$ 和 $\tilde{\boldsymbol{d}}$，立即可得(3.5)式。 ▮

定理 3.13　设 B 是 p 阶正定矩阵，$\boldsymbol{d}$ 为 p 维向量，对任意 p 维向量 $\boldsymbol{x}$ 下式成立

$$\max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{(\boldsymbol{x}'\boldsymbol{d})^2}{\boldsymbol{x}'B\boldsymbol{x}} = \boldsymbol{d}'B^{-1}\boldsymbol{d} \tag{3.6}$$

且当 $\boldsymbol{x} = cB^{-1}\boldsymbol{d}$ 时达到最大值 $\boldsymbol{d}'B^{-1}\boldsymbol{d}$ ($c \neq 0$ 常数)

证明　因为 B 是 p 阶正定阵，由引理 3.2 知道对任一 p 维向

量 $x \neq 0$, $x'Bx > 0$ 成立以及

$$\frac{(x'd)^2}{x'Bx} \leqslant d'B^{-1}d$$

且当 $x = cB^{-1}d$ 时等号成立,定理得证. ▌

定理 3.14 设 B 是 p 阶对称阵, $\lambda_i = \mathrm{ch}_i(B)$ 是 B 的第 i 大的特征根, e_i 是相应于 λ_i 的 B 的标准化特征向量 $(i = 1, \cdots, p)$, x 为任一非零 p 维向量,那么有

1)
$$\lambda_p \leqslant \frac{x'Bx}{x'x} \leqslant \lambda_1 \tag{3.7}$$

右边等号当 $x = ce_1$ 时成立,左边等号当 $x = ce_p$ 时成立,这里 c 是非零常数.

2) 记 $\mathscr{L}_2 = \mathscr{L}(e_{r+1}, \cdots, e_p)$, 则

$$\max_{\substack{x \neq 0 \\ x \in \mathscr{L}_2}} \frac{x'Bx}{x'x} = \lambda_{r+1} \tag{3.8}$$

当 $x = c \cdot e_{r+1}$ 时达到最大值,这里 c 非零常数.

证明 1) 记 $P = (e_1, \cdots, e_p)$, 它是正交阵,由定理 3.1 知

$$B = P\mathrm{diag}(\lambda_1, \cdots, \lambda_p)P'.$$

令 $y = P'x/\sqrt{x'x} = (y_1, \cdots, y_p)'$, 那么 $y'y = 1$ 以及

$$\frac{x'Bx}{x'x} = y'\mathrm{diag}(\lambda_1, \cdots, \lambda_p) \cdot y = \sum_{i=1}^{p} \lambda_i y_i^2$$

于是

$$\lambda_p = \lambda_p \sum_{i=1}^{p} y_i^2 \leqslant \sum_{i=1}^{p} \lambda_i y_i^2 \leqslant \lambda_1 \sum_{i=1}^{p} y_i^2 = \lambda_1$$

右边等式当 $y = (1, 0, \cdots, 0)'$ 时成立,从而

$$x/\sqrt{x'x} = Py = (e_1, \cdots, e_p)\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = e_1$$

即 $x = ce_1$。左边等式可类似证明。3.7 式得证。

2) 如果 $x \in \mathscr{L}(e_{r+1}, \cdots, e_p)$，则 $e_i'x = 0$ $(i = 1, \cdots, r)$。因此

$$y = P'x/\sqrt{x'x} = (0, \cdots, 0, y_{r+1}, \cdots, y_p),$$

$$y'y = \sum_{i=r+1}^{p} y_i^2 = 1$$

$$\frac{x'Bx}{x'x} = \sum_{i=r+1}^{p} \lambda_i y_i^2 \leqslant \lambda_{r+1} \sum_{i=r+1}^{p} y_i^2 = \lambda_{r+1}$$

而等号当 $y = (0, \cdots, 0, 1, 0, \cdots, 0)$时成立，其中 1 在 $r+1$ 位置上，同上可知此时 $x = ce_{r+1}$，(3.8)式得证。 ▮

推论 3.2 设 A 为 p 阶对称阵，B 为 p 阶正定矩阵，λ_i 是 A 相对于 B 的第 i 大特征根，e_i 是相应于 λ_i 的 A 相对于 B 的特征向量，那么对任一 p 维非零向量 x 有

$$\lambda_p \leqslant \frac{x'Ax}{x'Bx} \leqslant \lambda_1$$

右边等号当 $x = ce_1$ 时成立，左边等号当 $x = ce_p$ 成立，这里 c 是非零常数。

证明　由定理 3.11 知道存在非奇异矩阵 R，使得 $B = RR'$，$A = RD_\lambda R'$，其中 D_λ 是由 A 相对于 B 的诸特征根 λ_i 构成的对角阵。令 $y = R'x$，则有

$$\frac{x'Ax}{x'Bx} = \frac{y'D_\lambda y}{y'y}$$

由定理 3.14 知道

$$\lambda_p \leqslant \frac{x'Ax}{x'Bx} \leqslant \lambda_1$$

且当 $y = c\tilde{e}_1$ 时右边等号成立，其中 $\tilde{e}_1$ 是 D_λ 相应于 λ_1 的特征向量，即 $(\lambda_1 I - D_\lambda)y = 0$ 的解。由

$$(\lambda_1 I - D_\lambda)y = 0 \longleftrightarrow R(\lambda_1 I - D_\lambda)R'x = 0 \longleftrightarrow (\lambda_1 B - A)x = 0$$

知此时相应 x 正是相应于 λ_1 的 A 相对于 B 的特征向量 e_1 也即当 $x = ce_1$ 时右边等号成立。其他类似可证。 ∎

§1.4 矩阵的微分及其变换的 Jacobian 行列式

在通常的高等数学或数学分析教程中我们讨论过对自变量分别为一维变数或 p 维向量时它的函数的微分表达形式。而在本节我们将把上述概念推广到自变量为一维变数，p 维向量或者矩阵，而因变量为一维变数，q 维向量或者矩阵的各种情况。同时讨论当自变量和因变量分别为相同维数的向量或者相同阶数的矩阵时有关变换的 Jacobian 行列式。这些知识对于"多元分析"教程的学习是十分重要的，尤其在导出随机矩阵的分布密度时起着关键作用。另外本节在推导有关的结论时运用了一些数学上的技巧，这些方法都是很有启发性的。

(一) 矩阵的微分

我们将按自变量的各种情况(一维变数，向量，矩阵)分别讨论有关函数，向量函数及矩阵函数的微分表达形式。

(1) 自变量是一维变数 x。

① 若 $y = (y_1, \cdots, y_p)'$ 是 x 的向量函数，则记

$$\frac{dy}{dx} = \left(\frac{dy_1}{dx}, \cdots, \frac{dy_p}{dx}\right)'$$

为 y 对 x 的导数向量，而

$$dy = \left(\frac{dy_1}{dx}dx, \cdots, \frac{dy_p}{dx}dx\right)'$$

表示向量 y 的微分，它仍是一向量。

② 若 $Y = F(x)$ 是 x 的矩阵函数，其中 $Y = (y_{ij})$ 是 $p \times q$ 阶矩阵，那么

$$\frac{dY}{dx} = \left(\frac{dy_{ij}}{dx}\right)_{p\times q}$$

是 Y 对 x 的导数矩阵，它仍是 $p \times q$ 阶矩阵。

而

$$dY = \left(\frac{dy_{ij}}{dx}dx\right)_{p\times q}$$

表示矩阵 $Y = F(x)$ 的微分矩阵.

(2) 自变量是 p 维向量 $x = (x_1, \cdots, x_p)'$

① 若 $y = f(x)$ 是 x 的一维函数，通过令 $x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_p$ 为常数对 x_i 求导可求得 y 关于 x_i 的偏导数，记为 $\frac{\partial f}{\partial x_i}$ $\left(或 \frac{\partial y}{\partial x_i}\right)$，此时 y 关于 x 的偏导数向量记为

$$\frac{\partial f}{\partial x} = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_p}\right)'$$

如果令

$$\frac{\partial f}{\partial x'} = \left(\frac{\partial f}{\partial x}\right)' \text{ 和 } dx = (dx_1, \cdots, dx_p)'$$

则 $y = f(x)$ 的微分为

$$df = \frac{\partial f}{\partial x'}dx = \sum_{i=1}^{p}\frac{\partial f}{\partial x_i}dx_i \tag{4.1}$$

② 若 $y = (y_1, \cdots, y_q)'$ 是向量 x 的 q 维向量函数，即 $y_i = f(x)$ $(i = 1, \cdots, q)$，简记 $y = f(x)$. 那么规定 y 关于 x 的诸偏导数构成的矩阵为

$$\frac{\partial(y)}{\partial(x)} = \frac{\partial y}{\partial x'} = \left(\frac{\partial y_i}{\partial x_j}\right)_{q\times p}$$

$$= \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_p} \\ \vdots & & \\ \frac{\partial y_q}{\partial x_1} & \cdots & \frac{\partial y_q}{\partial x_p} \end{pmatrix}$$

若记 $dy = (dy_1, \cdots, dy_q)'$，它是一个向量，那么 y 关于 x 的微

分为

$$dy=\begin{pmatrix} dy_1 \\ \vdots \\ dy_q \end{pmatrix}=\begin{pmatrix} \frac{\partial y_1}{\partial x'}dx \\ \vdots \\ \frac{\partial y_q}{\partial x'}dx \end{pmatrix}=\frac{\partial(y)}{\partial(x)}dx.$$

③ 若 $Y=(y_{ij})$ 是 $n\times m$ 矩阵，而 $y_{ij}=f_{ij}(x)$，称 Y 是向量 x 的矩阵函数，记为 $Y=F(x)$. 那么类似上面 Y 关于 x 的微分仍是一个 $n\times m$ 矩阵

$$dY=dF(x)=\left(\frac{\partial f_{ij}(x)}{\partial x'}dx\right)_{n\times m}$$

我们记

$$\frac{\partial F}{\partial x_k}=\left(\frac{\partial f_{ij}(x)}{\partial x_k}\right)_{n\times m}\qquad(k=1,\cdots,p)$$

为 F 关于 x_k 的偏导数矩阵，那么 dF 还可表示为

$$dF=\sum_{k=1}^{p}\frac{\partial F}{\partial x_k}dx_k$$

(3) 自变量是 $p\times q$ 阶矩阵 $X=(x_{ij})$.

① $y=f(X)$ 是 X 的一维数值函数，那么偏导数矩阵为

$$\frac{\partial f}{\partial X}=\left(\frac{\partial f}{\partial x_{ij}}\right)_{p\times q}$$

仍为相同阶数的矩阵。 若记 $dX=(dx_{ij})$ 那么 y 的微分可写成

$$dy=\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\partial f}{\partial x_{ij}}dx_{ij}=\operatorname{tr}\left[\left(\frac{\partial f}{\partial X}\right)'dX\right]\qquad(4.2)$$

若自变量 X 是 $p\times p$ 阶对称矩阵，那么自变量只有 x_{ij} $(i\geqslant j, i,j=1,\cdots,p)$，为了能与非对称矩阵同样处理，我们可作如下技术处理：记

$$\frac{\partial f}{\partial X}=\left(\varepsilon_{ij}\frac{\partial f}{\partial x_{ij}}\right)$$

其中
$$\varepsilon_{ij}=\begin{cases}1, & i=j\\ \frac{1}{2}, & i\neq j\end{cases}$$

即 $\frac{\partial f}{\partial x_{ii}}$ 仍同往常放在对角元素上，而 $\frac{\partial f}{\partial x_{ij}}$ $(i\neq j)$ 各放一半在第 (i,j) 和 (j,i) 的位置上，而且继续记
$$dX=(dx_{ij})_{p\times p}$$
只是现在 $dx_{ij}=dx_{ji}$，那么
$$df=\sum_{i=1}^{p}\sum_{j=1}^{i}\frac{\partial f}{\partial x_{ij}}dx_{ij}=\operatorname{tr}\left[\left(\frac{\partial f}{\partial X}\right)'dX\right]$$
仍适用。

② 若向量 $y=(y_1,\cdots,y_n)'$ 是矩阵变数 X 的向量函数，即 $y_i=f_i(X)$ $(i=1,\cdots,n)$ 那么
$$dy=\begin{pmatrix}dy_1\\ \vdots\\ dy_n\end{pmatrix}=\begin{pmatrix}\operatorname{tr}\left(\frac{\partial f_1}{\partial X}\right)'dX\\ \vdots\\ \operatorname{tr}\left(\frac{\partial f_n}{\partial X}\right)'dX\end{pmatrix}$$
仍为一向量

③ 若 $n\times m$ 阶矩阵 Y 是 X 的一个矩阵函数，即 $y_{ij}=f_{ij}(X)$ $(i=1,\cdots,m,\ j=1,\cdots,n)$. 那么 Y 关于 X 的微分仍为一个相同阶数矩阵
$$dY=\left(\operatorname{tr}\left[\left(\frac{\partial f_{ij}}{\partial X}\right)'dX\right]\right)_{n\times m}$$

下面以例子说明。

例 4.1 设 $F=GH$，其中 $F=(f_{ij})$，$G=(g_{ij})$，$H=(h_{ij})$ 分别是向量变数 $x=(x_1,\cdots,x_p)'$ 的矩阵函数，求 dF。

解 由 $F=GH$ 知，$f_{ij}=\sum_{s}g_{is}h_{sj}$。而

$$\frac{\partial f_{ij}}{\partial x_k}=\sum_t \frac{\partial g_{it}}{\partial x_k} h_{tj}+\sum_t g_{it} \frac{\partial h_{tj}}{\partial x_k}$$

于是

$$\frac{\partial F}{\partial x_k}=\frac{\partial G}{\partial x_k}\cdot H+G\cdot\frac{\partial H}{\partial x_k}$$

由此可得

$$dF=\sum_k \frac{\partial F}{\partial x_k} dx_k=(dG)\cdot H+G\cdot(dH).$$

与通常函数乘积求微分有相同形式。▌

例 4.2 已知 $X=(x_{ij})$ 是一对称非奇异方阵，求 $F=X^{-1}$ 的微分。

解 由于 $XF=I$，所以

$$dX\cdot F+XdF=0$$

$$dF=-X^{-1}\cdot dX\cdot X^{-1}$$ ▌

例 4.3 设 $f(X)=a'X^{-1}a$，a 是常数向量，$X=(x_{ij})$ 是对称非奇异方阵，求 $\frac{\partial f}{\partial X}$

解 首先从定义容易验证 $d\mathrm{tr}(X)=\mathrm{tr}(dX)$。于是

$$\begin{aligned} df(X)&=d[\mathrm{tr}(a'X^{-1}a)]=\mathrm{tr}[d(a'X^{-1}a)]\\ &=\mathrm{tr}[a'\cdot dX^{-1}\cdot a]=-\mathrm{tr}(a'X^{-1}\cdot dX\cdot X^{-1}\cdot a)\\ &=\mathrm{tr}[(-X^{-1}\cdot a\cdot a'\cdot X^{-1})'dX]\end{aligned}$$

这里利用 X 对称性。由 (4.2)式得到

$$\frac{\partial f}{\partial X}=-X^{-1}\cdot a\cdot a'\cdot X^{-1}$$ ▌

注意：求矩阵变数的函数的偏导数矩阵常用例 4.3 方法，即先求出 $df=\mathrm{tr}(A'dX)$，那么 $\frac{\partial f}{\partial X}=A$.

例 4.4 求证 $d|X|=|X|\mathrm{tr}(X^{-1}dX)$.

证明 由定理 1.2 知，若记 $X^{-1}=(x^{ij})$，那么 x_{ij} 所对应的

代数余子式正是 $A(i,j)=|X|\cdot x^{ji}$，由于

$$|X|=\sum_i x_{ij}A(i,j)$$

所以

$$\frac{\partial |X|}{\partial x_{ij}}=A(i,j)=|X|\cdot x^{ji}$$

于是

$$d|X|=\sum_i\sum_j |X|\cdot x^{ji}dx_{ij}=|X|\cdot \operatorname{tr}(X^{-1}dX). \quad \blacksquare$$

读者可自行验算下列微分公式，其中 X, Y 为矩阵变数，A，B 是常数矩阵.

1) $d(\operatorname{tr}X)=\operatorname{tr}(dX)$.

2) $d(AX)=AdX$.

3) $d(XY)=dX\cdot Y+X\cdot dY$.

4) $d(X\otimes Y)=dX\otimes Y+X\otimes dY$.

5) $d(X'BX)=(dX')\cdot BX+X'B\cdot(dX)$.

6) $d(\operatorname{tr}X'AXB)=\operatorname{tr}[(B'X'A'+BX'A)dX]$.

7) $d(\ln|X|)=\operatorname{tr}(X^{-1}dX)$.

在上述 2)及 5)式中若令 $X=x$ 是一向量，则可推出

$$\frac{\partial(Ax)}{\partial x'}=A \tag{4.3}$$

以及

$$\frac{\partial x'Bx}{\partial x}=(B+B')x$$

特别当 B 是对称矩阵时有

$$\frac{\partial(x'Bx)}{\partial x}=2Bx \tag{4.4}$$

（二） 矩阵变换的 Jacobian 行列式

下面研究变换的 Jacobian 行列式，在多元分布的推导过程

中常常在积分号下作变换，这时必须计算它的 Jacobian 行列式.

考虑积分

$$\int_{\Omega} g(\boldsymbol{x})d\boldsymbol{x} \qquad \Omega \in R^p$$

其中 $\boldsymbol{x}=(x_1,\cdots,x_p)'$. 作变换 $\boldsymbol{y}=f(\boldsymbol{x})$. 即 $Y=(y_1,\cdots,y_p)'$. $y_i=f_i(x)$ $(i=1,\cdots,p)$. (当逆变换存在时记 $\boldsymbol{x}=f^{-1}(\boldsymbol{y})$). 由微积分知识知道上面积分可化为

$$\int_T g(f^{-1}(\boldsymbol{y}))J(\boldsymbol{x}\to\boldsymbol{y})d\boldsymbol{y} \tag{4.5}$$

其中积分区域 $T=\{\boldsymbol{y}:\boldsymbol{y}=f(\boldsymbol{x}),\boldsymbol{x}\in\Omega\}$

$$J(\boldsymbol{x}\to\boldsymbol{y})=\left|\frac{\partial(\boldsymbol{x})}{\partial(\boldsymbol{y})}\right|_+ \quad (|A|_+ \text{ 表示行列式} |A| \text{绝对值})$$

$J(\boldsymbol{x}\to\boldsymbol{y})$ 称为变换的 Jacobian 行列式.

注意到这里自变量 $\boldsymbol{x}$ 和因变量 $\boldsymbol{y}$ 有相同的维数.

若 $\boldsymbol{y}=f(\boldsymbol{x})$, 计算 $J(\boldsymbol{y}\to\boldsymbol{x})$ 常用下列方法: 由于

$$d\boldsymbol{y}=\frac{\partial(\boldsymbol{y})}{\partial(\boldsymbol{x})}d\boldsymbol{x}$$

我们不妨把 $d\boldsymbol{y}$ 看成 $d\boldsymbol{x}$ 的线性函数，由矩阵微分性质 2)知它相应的 Jacobian 行列式正是 $\left|\frac{\partial(\boldsymbol{y})}{\partial(\boldsymbol{x})}\right|_+$, 于是有

$$J(\boldsymbol{y}\to\boldsymbol{x})=J(d\boldsymbol{y}\to d\boldsymbol{x})=\{J[\boldsymbol{x}\to\boldsymbol{y}]\}^{-1} \tag{4.6}$$

如果自变量 X 和因变量 Y 都是 $p\times q$ 阶矩阵变数，则通常用

$$J(Y\to X)=\left|\frac{\partial(\mathrm{vec}Y)}{\partial(\mathrm{vec}X)}\right|_+ \tag{4.7}$$

来定义从 Y 到 X 的变换的 Jacobian 行列式.

*)**定理 4.1** 对于 $i=1,\cdots,n$, 规定下列条件变换

$$\begin{cases} y_1=f_1(x_1,x_2,\cdots,x_n) \\ y_2=f_2(y_1,x_2,\cdots,x_n) \\ \cdots\cdots\cdots \\ y_n=f_n(y_1,y_2\cdots,x_n) \end{cases}$$

其中 $\boldsymbol{x}_i$ 和 $\boldsymbol{y}_i$ 都是 q_i 维向量，且 $\boldsymbol{y}_i$ 表达式中不显含 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{i-1}$ 那么有

$$J(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n\to\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)=\prod_{i=1}^{n}J(\boldsymbol{y}_i\to\boldsymbol{x}_i)\tag{4.8}$$

证明　对 $\boldsymbol{y}_i$ 求微分得到

$$d\boldsymbol{y}_i=\sum_{j=1}^{i-1}\frac{\partial f_i}{\partial\boldsymbol{y}_j'}d\boldsymbol{y}_j+\sum_{j=i}^{n}\frac{\partial f_i}{\partial\boldsymbol{x}_j'}d\boldsymbol{x}_j\ (i=1,\cdots,n)\tag{4.9}$$

若记 $\boldsymbol{x}=(\boldsymbol{x}_1',\cdots,\boldsymbol{x}_n')'$，$\boldsymbol{y}=(\boldsymbol{y}_1',\cdots,\boldsymbol{y}_n')'$，它们都是 $\sum_{i=1}^{n}q_i$ 维向量

$$d\boldsymbol{x}=((d\boldsymbol{x}_1)',\cdots,(d\boldsymbol{x}_n)')'\quad d\boldsymbol{y}=((d\boldsymbol{y}_1)',\cdots,(d\boldsymbol{y}_n)')'.$$

并记

$$\boldsymbol{T}_{ij}=\begin{cases}\dfrac{\partial f_i}{\partial\boldsymbol{x}_j'} & (i\leqslant j)\\ 0 & (i>j).\end{cases}$$

$$\boldsymbol{G}_{ij}=\begin{cases}-\dfrac{\partial f_i}{\partial\boldsymbol{y}_j'} & (i>j)\\ I_{qi} & (i=j)\\ 0 & (i<j).\end{cases}$$

$$\boldsymbol{T}=\begin{pmatrix}\boldsymbol{T}_{11}&\cdots&\boldsymbol{T}_{1n}\\ &\ddots&\vdots\\ 0&&\boldsymbol{T}_{nn}\end{pmatrix}\quad \boldsymbol{G}=\begin{pmatrix}\boldsymbol{G}_{11}&&0\\ \vdots&\ddots&\\ \boldsymbol{G}_{n1}&\cdots&\boldsymbol{G}_{nn}\end{pmatrix}$$

那么上述方程组(4.9)可改写成

$$\boldsymbol{G}d\boldsymbol{y}=\boldsymbol{T}d\boldsymbol{x}$$

注意到 G 是对角元素全为 1 的下三角矩阵，非奇异且 $|G|=1$，所以

$$d\boldsymbol{y}=\boldsymbol{G}^{-1}\boldsymbol{T}d\boldsymbol{x}$$

$$J(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n\to\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)=J(\boldsymbol{y}\to\boldsymbol{x})$$
$$=|\boldsymbol{G}^{-1}\boldsymbol{T}|_+=|\boldsymbol{G}^{-1}|_+\cdot|\boldsymbol{T}|_+=|\boldsymbol{T}|_+$$

$$= \prod_{i=1}^{n} |T_{ii}|_+ = \prod_{i=1}^{n} J(y_i \to x_i). \blacksquare$$

定理 4.2 设 $\boldsymbol{y} = f(\boldsymbol{x})$, $\boldsymbol{x} = g(\boldsymbol{z})$, 其中 $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ 都是 n 维向量，那么

$$J(\boldsymbol{y} \to \boldsymbol{z}) = J(\boldsymbol{y} \to \boldsymbol{x}) \cdot J(\boldsymbol{x} \to \boldsymbol{z}).$$

证明　通过取微分知道

$$d\boldsymbol{y} = \frac{\partial(f)}{\partial(\boldsymbol{x})} d\boldsymbol{x} \qquad d\boldsymbol{x} = \frac{\partial(g)}{\partial(\boldsymbol{z})} d\boldsymbol{z}$$

所以

$$d\boldsymbol{y} = \frac{\partial(f)}{\partial(\boldsymbol{x})} \cdot \frac{\partial(g)}{\partial(\boldsymbol{z})} d\boldsymbol{z}$$

因此

$$J(\boldsymbol{y} \to \boldsymbol{z}) = J(d\boldsymbol{y} \to d\boldsymbol{z}) = \left|\frac{\partial(f)}{\partial(\boldsymbol{x})}\right|_+ \cdot \left|\frac{\partial(g)}{\partial(\boldsymbol{z})}\right|_+$$

$$= J(\boldsymbol{y} \to \boldsymbol{x}) \cdot J(\boldsymbol{x} \to \boldsymbol{z}). \quad \blacksquare$$

上述定理 4.1 和定理 4.2 在求变换 Jacobian 行列式时很有用处，往往比由(4.7)式直接计算要方便得多，下面举出一些实例。

例 4.5　证明下列球坐标变换

$$\begin{cases} x_1 = r\sin\theta_1 \\ x_2 = r\cos\theta_1\sin\theta_2 \\ \vdots \\ x_i = r\cos\theta_1\cdots,\cos\theta_{i-1}\sin\theta_i \\ \vdots \\ x_p = r\cos\theta_1\cdots,\cos\theta_{p-1}\cos\theta_{p-1} \end{cases} \tag{4.10}$$

的变换 Jacobian 行列式为

$$J(x_1,\cdots,x_p \to r,\theta_1,\cdots,\theta_{p-1}) = r^{p-1}\prod_{i=1}^{p-1}(\cos\theta_i)^{p-i-1}$$

证明　为了利用定理 4.1，把上述变换变形为

$$\begin{cases} x_1 = r\sin\theta_1 = f_1(r,\theta_1,\cdots,\theta_{p-1}) \\ x_2 = x_1\mathrm{ctg}\theta_1\sin\theta_2 = f_2(x_1,\theta_1,\cdots,\theta_{p-1}) \\ \vdots \\ x_i = x_{i-1}\mathrm{ctg}\theta_{i-1}\sin\theta_i = f_i(x_1,\cdots,x_{i-1},\theta_{i-1},\cdots,\theta_{p-1}) \\ \vdots \\ x_p = x_{p-1}\mathrm{ctg}\theta_{p-1} = f_p(x_1,\cdots,x_{p-1},\theta_{p-1}) \end{cases}$$

于是

$$J(x_1,\cdots,x_p \to r,\theta_1,\cdots,\theta_{p-1}) = J(x_1 \to r)\cdot\prod_{i=2}^{p-1} J(x_i \to \theta_{i-1})$$

$$\cdot J(x_p \to \theta_{p-1}) = \sin\theta_1\cdot\prod_{i=2}^{p-1} x_{i-1}(\sin\theta_{i-1})^{-2}\sin\theta_i\cdot x_{p-1}$$

$$\cdot(\sin\theta_{p-1})^{-2}$$

将(4.10)式代入，消去诸 $\sin\theta_i$，得到

$$\begin{aligned} &J(x_1,\cdots,x_p \to r,\theta_1,\cdots,\theta_{p-1}) \\ &= r^{p-1}(\cos\theta_1)^{p-2}(\cos\theta_2)^{p-3}\cdots(\cos\theta_{p-2}) \\ &= r^{p-1}\prod_{i=1}^{p-1}(\cos\theta_i)^{p-i-1}. \quad \blacksquare \end{aligned}$$

推论 4.1 令 Y 和 X 都是 $n\times p$ 阶矩阵变数，并满足关系式 $Y = A\cdot X\cdot B$，其中 A 是 n 阶非奇异方阵，B 是 p 阶非奇异方阵，那么

$$J(Y \to X) = |A|_+^p\cdot|B|_+^n$$

证明 令 $Y = (y_1,\cdots,y_p)$ $X = (x_1,\cdots,x_p)$，若 $Y = AX$ 则

$$J(AX \to X) = \prod_{i=1}^{p} J(y_i \to x_i) = \prod_{i=1}^{p}|A|_+ = |A|_+^p.$$

同理若 $Y = XB$，只须考虑行向量，则有

$$J(XB \to X) = |B|_+^n.$$

于是由定理 4.2 求得

$$J(Y\to X)=J(AXB\to XB)\cdot J(XB\to X)=|A|_+^p\cdot|B|_+^n. \blacksquare$$

注：推论 4.1 也可利用定理 1.1 证明，即 $Y=AXB$ 可改写为 $\mathrm{vec}Y=(A\otimes B')\mathrm{vec}X$，按(4.7)

$$J(Y\to X)=\left|\frac{\partial(\mathrm{vec}Y)}{\partial(\mathrm{vec}X)}\right|_+=|A\otimes B'|_+=|A|_+^p|B|_+^n.$$

*)**定理 4.3** 令 V 和 W 是两个 n 阶对称矩阵变数且满足

$$V=AWA'$$

其中 A 是 n 阶非奇异矩阵，那么

$$J(V\to W)=|AA'|^{\frac{1}{2}(n+1)}=|A|_+^{n+1}.$$

证明 由于交换 A 的诸行(列)时，只须把 V 的诸行(列)作相应变换而 V 和 W 的关系不变。又因为 A 非奇异，不妨设 A 诸 i 阶主子行列式都不为 0，由定理 1.7 知 A 可分解为 $A=TU$，其中 T 为下三角矩阵，U 为上三角矩阵，因此

$$V=TU\cdot W\cdot U'T'=TW_1T' \quad (\text{其中 } W_1=UWU') \tag{4.11}$$

从而

$$J(V\to W)=J(V\to W_1)\cdot J(W_1\to W).$$

不妨设 A 为下三角矩阵 T，我们用归纳法证明.

当阶数为 1 时，定理显然成立。设阶数为 $n-1$ 时定理也成立，那么当阶数为 n 时，把变换写成分块形式：

$$V=\begin{pmatrix}V_{11} & b\\ b' & v\end{pmatrix}=\begin{pmatrix}T_{11} & 0\\ t' & t_1\end{pmatrix}\begin{pmatrix}W_{11} & c\\ c' & w\end{pmatrix}\begin{pmatrix}T'_{11} & t\\ 0 & t_1\end{pmatrix}$$

其中 b 和 t 是 $n-1$ 维向量，T_{11} 是 $n-1$ 阶下三角矩阵，v，t_1，w 是常数。把上式改写成

$$\begin{cases}V_{11}=T_{11}W_{11}T'_{11}\\ b=T_{11}W_{11}t+T_{11}\cdot c\cdot t_1\\ v=t_1^2\cdot w+2t_1\cdot c't+t'W_{11}t\end{cases}$$

正是条件变换式，按定理 4.1 有

$$J(V\to W)=J(V_{11}\to W_{11})\cdot J(b\to c)\cdot J(v\to w)$$

容易计算得到

$$J(b\to c)=|T_{11}|_+|t_1|_+^{n-1}\qquad J(v\to w)=t_1^n$$

由归纳法

$$J(V_{11}\to W_{11})=|T_{11}|_+^n.$$

所以有

$$J(V\to W)=|T_{11}|_+^n\cdot|T_{11}|_+\cdot|t_1|_+^{n+1}=|T|_+^{n+1}$$

类似地当 A 为上三角矩阵 U 时，则有

$$J(V\to W)=|U|_+^{n+1}$$

按(4.11)式立即得到

$$\begin{aligned}J(V\to W)&=J(V\to W_1)\cdot J(W_1\to W)\\&=|T|_+^{n+1}\cdot|U|_+^{n+1}=|A|_+^{n+1}.\end{aligned}\quad ■$$

*)**定理 4.4** 设 S 是 p 阶正定矩阵，且 $S=T\cdot T'$，其中 T 是具有正对角元素的下三角矩阵，那么

$$J(S\to T)=2^p\prod_{i=1}^{p}t_{ii}^{p-i+1}\tag{4.12}$$

证明　利用归纳法．当阶数为 1 时 $s_{11}=t_{11}^2$，于是

$$J(s_{11}\to t_{11})=2t_{11}$$

(4.12)式成立．设阶数为 $p-1$ 时定理成立．当阶数为 p 时，S 是 p 阶正定矩阵，由定理 3.12 知道存在对角元素为正的下三角矩阵 T 使得 $S=T\cdot T'$．将它用分块形式表示

$$\begin{pmatrix}S_{11}&u\\u'&s_{pp}\end{pmatrix}=\begin{pmatrix}T_{11}&0\\v'&t_{pp}\end{pmatrix}\begin{pmatrix}T'_{11}&v\\0&t_{pp}\end{pmatrix}$$

其中 u,v 是 $p-1$ 维向量，s_{pp}，t_{pp} 是一维变数，T_{11} 是对角元素为正的 $p-1$ 阶下三角矩阵，即

$$\begin{cases}S_{11}=T_{11}\cdot T'_{11}\\u=T_{11}v\\s_{pp}=v'v+t_{pp}^2\end{cases}$$

是条件变换式，按定理 4.1，有

$$J(S\to T)=J(S_{11}\to T_{11})\cdot J(u\to v)\cdot J(s_{pp}\to t_{pp})$$

容易求得

$$J(u \to v) = |T_{11}| = \prod_{i=1}^{p-1} t_{ii}, \quad J(s_{pp} \to t_{pp}) = 2t_{pp}$$

由归纳法

$$J(S_{11} \to T_{11}) = 2^{p-1} \prod_{i=1}^{p-1} t_{ii}^{p-i}$$

于是

$$J(S \to T) = 2^{p-1} \prod_{i=1}^{p-1} t_{ii}^{p-i} \prod_{i=1}^{p-1} t_{ii} \cdot 2t_{pp}$$

$$= 2^p \cdot \prod_{i=1}^{p} t_{ii}^{p-i+1}. \quad ▮$$

*) **定理 4.5** 设 V 是一个 p 阶正定矩阵变数，V^{-1} 是它的逆阵，则

$$J(V \to V^{-1}) = |V|^{p+1}$$

证明 记 $W = V^{-1}$，则 $V \cdot W = I$. 所以 $VdW = -dV \cdot W$ 即

$$dV = -VdW \cdot V$$

利用(4.6)式和定理 4.3 可得到

$$J(V \to W) = J(dV \to dW) = |V|^{p+1}$$

注意若 V 是 p 阶非奇异矩阵，那么 $J(V \to V^{-1}) = |V|^{2p}$（见本章习题 16）

习 题 一

1.1 计算行列式：

$$a) \begin{vmatrix} 1 & a_1 & \cdots & a_1^{n-1} \\ 1 & a_2 & \cdots & a_2^{n-1} \\ \vdots & & & \\ 1 & a_n & \cdots & a_n^{n-1} \end{vmatrix}_{n\times n}; \qquad b) \begin{vmatrix} 1 & \rho & \cdots & \rho & \rho \\ \rho & 1 & \cdots & \rho & \rho \\ \vdots & \vdots & & \vdots & \vdots \\ \rho & \rho & \cdots & \rho & 1 \end{vmatrix}_{n\times n}.$$

1.2 求下列方阵的逆矩阵:

a) 设 $D_p = \mathrm{diag}(p_1, \cdots, p_n)$, $p = (p_1, \cdots, p_n)'$,

$$\left(\sum_{i=1}^{n} p_i < 1,\ p_i > 0,\ i = 1, \cdots, n\right)$$

$$V = D_p - p \cdot p',$$

求 V^{-1}。

b) $V = (v_{ij})_{n \times n}$, $v_{ij} = \begin{cases} a, & i = j, \\ \rho, & i \neq j, \end{cases}$

求 V^{-1}。

提示: 利用(2.7)式。

1.3 证明: a) $\mathrm{rk}(AB) = \mathrm{rk}(A'AB) = \mathrm{rk}(ABB')$;

b) $|\mathrm{rk}(A) - \mathrm{rk}(B)| \leqslant \mathrm{rk}(A + B) \leqslant \mathrm{rk}(A) + \mathrm{rk}(B)$.

1.4 设 A 是秩为 r 的 $n \times m$ 阶矩阵，则存在秩为 r 的 $n \times r$ 阶矩阵 F 和 $r \times m$ 阶矩阵 G，使得 $A = F \cdot G$. 记 $B_0 = G'(GG')^{-1}(F'F)^{-1}F'$，证明 B_0 满足加号逆的四个条件。

1.5 证明: 若 P_1, P_2 是正交阵，则 $(P_1AP_2)^+ = P_2'A^+P_1'$.

1.6 设 A 是反对称矩阵，即 A 满足 $A' = -A$. 求证

a) A^2 是对称阵;

b) 奇数阶的反对称阵行列式为 0;

c) 反对称阵的特征根必为零或者纯虚数。

*)1.7 求下列矩阵的加号逆:

a) $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$; b) $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

1.8 设 $B = I_n - \frac{1}{n} J_n$, $C_i = I_{n_i} - \frac{1}{n_i} J_{n_i}$ $\left(i = 1, \cdots, k,\ \sum_{i=1}^{k} n_i = n\right)$.

$C = \mathrm{diag}(C_1, \cdots, C_k)$, $D = B - C$,

其中 J_n 表示元素都是 1 的 $n \times n$ 阶矩阵. 证明: B, C, D 都是对称幂等阵。

1.9 若 A 是 $p \times q$ 阶矩阵，B 是 $q \times p$ 阶矩阵. 证明: AB 和 BA 有相同的非零特征根。

1.10

a) 证明 如果 A^{-1}, B^{-1} 存在,则 $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$;

b) 证明 $\mathrm{tr}(A\otimes B) = \mathrm{tr}A \cdot \mathrm{tr}B$;

c) 设 A, B 为方阵, α 是 A 相应于 A 的特征根 λ 的特征向量, v 是 B 相应于 B 的特征根 μ 的特征向量,证明: $\alpha\otimes v$ 是 $A\otimes B$ 相应于它的特征根 $\lambda\mu$ 的特征向量.

1.11 已知 X 是 $n\times m$ 阶矩阵变数,

a) $f(X) = \mathrm{tr}(X'AXB)$, 求 $\partial f/\partial X$;

b) $f(X) = \ln|X|$, 求 $\partial f/\partial X$.

1.12 设 A 和 B 是两个 $n\times p$ 阶矩阵, 且 $\mathscr{L}(A)\subset\mathscr{L}(B)$, $P_A = A(A'A)^- A'$, $P_B = B(B'B)^- B'$ 分别是从 R^n 到 $\mathscr{L}(A)$ 和 $\mathscr{L}(B)$ 的投影矩阵,证明 $P_A\cdot P_B = P_B\cdot P_A = P_A$.

提示: 若 $\mathscr{L}(A)\subset\mathscr{L}(B)$, 则存在 $p\times p$ 阶矩阵 D 使得 $A = BD$.

1.13 证明若一个方阵既是上三角矩阵又是正交矩阵,那么它一定是单位阵。

*)1.14 如果方阵 V 和 W 都是 $n\times n$ 阶反对称矩阵变数, 且 $V = AWA'$ (其中 A 是 $n\times n$ 阶非奇异矩阵)证明:

$$J(V\to W) = |A|^{n-1}.$$

提示: 证明思路参考定理 4.3.

*)1.15 如果 $S = T\cdot\Lambda\cdot T'$, 其中 T 是对角元素全为 1 的下三角矩阵,又 $\Lambda = \mathrm{diag}(\lambda_1,\cdots,\lambda_p)$. 证明:

$$J(S\to T,\Lambda) = \prod_{i=1}^{p}\lambda_i^{p-i}$$

提示同 1.14 题.

*)1.16 已知 V 是 $p\times p$ 阶非奇异矩阵, V^{-1} 是它的逆阵,求证:

$$J(V\to V^{-1}) = |V|^{2p}$$

提示: 利用(4.6)式和推论 4.1.

第二章　多元正态分布

从本章起，我们将逐步介绍多维随机变数的统计规律．在通常的数理统计教材，为了强调随机变量与普通变量之间的差异，往往用不同的印刷符号来表示．而在本教程中认为读者已具有普通的数理统计知识，不再对随机变量和普通变量加以区别，根据上下文，读者是能够自行识别的．例如：EX 表示随机矩阵 X 的数学期望，而密度 $f(X)=c\cdot\exp\left\{-\frac{1}{2}\mathrm{tr}XX'\right\}$ 中 X 则是普通的矩阵变数．今后总是用大写字母表示矩阵变数，以小写黑体字母表示向量，而小写普通字母则表示一维变数．

在多元统计分析中，多元正态分布占有很重要的地位，这一方面是由于自然界中有许多随机现象确实服从多元正态分布，或者根据中心极限定理可推出许多实际问题中出现的多元随机变量渐近地服从多元正态分布．另一方面是因为有关多元正态分布的理论研究已经相当深入，有一整套的统计推断方法，这些方法和研究结果在实践中已被证明是十分有效的．为了更好地讨论多元正态分布，我们将在本章第一节中先回顾一下多元随机向量的分布及其性质，然后对多元正态分布本身有关的统计特性进行比较深入的讨论．而在以后的几章中将进一步研究与它密切相关的其他多元分布并讨论有关的统计推断问题．

§2.1　多元随机向量及其分布

（一）随机向量及其分布

有些随机现象，不能仅用一个随机变量来描述，而要同时使用几个随机变量才能全面描述．例如某种地质样品，我们就必须了解其中铜、锌、铅等元素的含量及它们之间的关系才能作出比较合

理的推断。这样我们就必须对随机向量进行研究。

设 $x_1, \cdots, x_p$ 是 p 个随机变量，由它们构成的向量 $x = (x_1, \cdots, x_p)'$ 称为 p 维随机向量。它的统计特性完全是由它的联合概率分布或分布函数所决定。当然我们也可以对随机向量的分量各自进行研究。但我们将看到，如果作为一个整体进行研究，不仅能了解各个分量的性质，而且可以考查各随机分量之间的统计联系，可以更深刻地揭示随机现象的本质。有时为了需要，常把许多随机变量整理成矩阵形式。如果一个 $n \times p$ 阶矩阵 X，它的每一项都是随机变量，则称 X 是**随机矩阵**。由于矩阵可以拉长成向量，故我们今后着重研究随机向量。

定义 1.1 设 $x = (x_1, \cdots, x_p)'$ 是 p 维随机向量，$y = (y_1, \cdots, y_p)'$ 是 p 维实数向量，称 p 元函数

$$F(y_1, \cdots, y_p) = P(x_1 < y_1, \cdots, x_p < y_p)$$

为 x 的联合分布函数。记作 $x \sim F$

给出了联合分布函数，那么 x 落在 R^p 中任一立方体 $I_{[a,b)}$ 中的概率可以通过 F 进行计算，其中 $a = (a_1, \cdots, a_p)'$，$b = (b_1, \cdots, b_p)'$，且 $a_i \leqslant b_i$ $(i = 1, \cdots, p)$，$I_{[a,b)} = \{(x_1, \cdots, x_p) \in R^p \quad a_1 \leqslant x_1 < b_1, \cdots, a_p \leqslant x_p < b_p\}$，即

$$P\{x \in I_{[a,b)}\} = \sum_{x \in A} F(x_1, \cdots, x_p) - \sum_{x \in B} F(x_1, \cdots, x_p) \triangleq \Delta_{[a,b)} F \quad (1.1)$$

其中 $A = \{(x_1, \cdots, x_p)$；其中只有偶数个 x_i 取 a_i，其余 x_i 取 $b_i\}$，$B = \{(x_1, \cdots, x_p)$；其中只有奇数个 x_i 取 a_i，其余 x_i 取 $b_i\}$。例如当 $p = 2$ 时

$$P\{a_1 \leqslant x_1 < b_1, a_2 \leqslant x_2 < b_2\} = F(b_1, b_2) + F(a_1, a_2) - F(a_1, b_2) - F(a_2, b_1)$$

定理 1.1 设 $F(x) = F(x_1, \cdots, x_p)$ 是 p 维分布函数，它具

有下列性质

1）对每个自变量 x_i，$F(x_1,\cdots,x_p)$ 是 x_i 的单调非降左连续函数（$i=1,2,\cdots,p$）.

2）$0\leqslant F(x_1,\cdots,x_p)\leqslant 1$.

3）对任意 $i(i=1,\cdots,p)$，

$$F(x_1,\cdots,x_{i-1},-\infty,x_{i+1},\cdots,x_p)=\lim_{x_i\to-\infty}F(x_1,\cdots,x_p)=0$$

以及 $$F(+\infty,\cdots,+\infty)=\lim_{\substack{x_i\to+\infty\\(i=1,\cdots,p)}}F(x_1,\cdots,x_p)=1.$$

4）$\Delta_{[a,b)}F\geqslant 0$（其中 a、b 满足(1.1)式中的要求）.

反之，若一 p 元函数 $F(x)$ 具有上述各条性质，它必为某一 p 维随机向量的分布函数. 定理详细证明可参考[35].

与一元随机变量一样，对于 p 维随机向量我们主要研究离散型和连续型两类.

定义 1.2 如果 p 维随机向量 $x=(x_1,\cdots,x_p)'$ 的每个分量都是一维离散型的，则 x 称为离散型的.

如果存在非负的 p 元函数 $f(x)$，使得 x 的分布函数 $F(x)$ 可以表示成

$$F(x_1,\cdots,x_p)=\int_{-\infty}^{x_p}\cdots\int_{-\infty}^{x_1}f(t_1,\cdots,t_p)dt_1\cdots dt_p \quad (1.2)$$

则 x 称为连续型的. 而 $f(x)$ 为 x 的分布密度函数，简称密度.

对于离散型随机向量，常用它的概率分布代替它的分布函数. 设 $x=(x_1,\cdots,x_p)'$ 的取值范围为点 $(a_1^{(i_1)},\cdots,a_p^{(i_p)})'$ 其中 $i_1,\cdots,i_p$ 可以是有限或者可数个，那么由

$$p(i_1,\cdots,i_p)=P(x_1=a_1^{(i_1)},\cdots,x_p=a_p^{(i_p)}) \quad (1.3)$$

全体组成了 x 的概率分布. 此时有

1）$p(i_1,\cdots,i_p)\geqslant 0$.

2）$\Sigma p(i_1,\cdots,i_p)=1$（求和是在 x 全体取值点上进行的）.

3）x 的分布函数

$F(y_1, \cdots, y_p) = \Sigma p(i_1, \cdots, i_p)$（求和是在区域 $A = \{(a_1^{(i_1)}, \cdots, a_p^{(i_p)}),\ a_1^{(i_1)} < y_1, \cdots, a_p^{(i_p)} < y_p\}$ 上进行的。

例 1.1 （多项分布）若随机向量 $\boldsymbol{x} = (x_1, \cdots, x_r)'$ 满足下列条件

i）$x_i \geqslant 0$ 取整数，且 $\sum_{i=1}^{r} x_i = n$；

ii）设 $\boldsymbol{a} = (a_1, \cdots, a_r)'\left(\text{其中 } a_i \text{ 为非负整数且 } \sum_{i=1}^{r} a_i = n\right)$，

$$P(x_1 = a_1, \cdots, x_r = a_r) = \frac{n!}{a_1! \cdots a_r!} p_1^{a_1} \cdots p_r^{a_r},$$

其中 $0 < p_i < 1$，且 $\sum_{i=1}^{r} p_i = 1$，则称随机向量 $\tilde{x} = (x_1, \cdots, x_{r-1})'$ 服从参数为 $(n, p_1, \cdots, p_{r-1})$ 的多项分布，记为 $\tilde{x} \sim PN(n, p_1, \cdots, p_{r-1})$。注意到 x_r 的取值完全由 $x_1, \cdots, x_{r-1}$ 和 n 决定的（例如当 $r = 2$ 时 x_1 的分布就是二项分布）。有时也记为 $\boldsymbol{x} \sim PN(n, p_1, \cdots, p_r)$。

对于连续型随机向量，通常研究它的分布密度。p 维随机向量 $\boldsymbol{x}$ 的密度 $f(\boldsymbol{x}) = f(x_1, \cdots, x_p)$ 具有下列性质：

i）$f(\boldsymbol{x}) \geqslant 0$，$\forall \boldsymbol{x} \in R^p$；

ii）$\int_{R^p} f(\boldsymbol{x}) d\boldsymbol{x} = 1$；

iii）对于 R^p 的任一集合 B(*) 有

$$P(\boldsymbol{x} \in B) = \int_B f(\boldsymbol{x}) d\boldsymbol{x};$$

iv）如果 $\boldsymbol{x}_0$ 是 $f(\boldsymbol{x})$ 的连续点，那么

$$f(\boldsymbol{x}_0) = \frac{\partial^p}{\partial x_1 \cdots \partial x_p} F(x_1, \cdots, x_p)\Big|_{\boldsymbol{x}=\boldsymbol{x}_0}$$

其中 $F(\boldsymbol{x})$ 是 $\boldsymbol{x}$ 的分布函数。

(*) 严格地说，B 应是 R^p 中的 Borel 可测集，以后提到的集合都有相同的说明，不再一一强调。

反之，如果一个 p 维非负可测函数 $f(\boldsymbol{x})$ 满足 i) 和 ii)，那么由

$$F(x_1,\cdots,x_p)=\int_{-\infty}^{x_p}\cdots\int_{-\infty}^{x_1}f(t_1,\cdots,t_p)dt_1\cdots dt_p$$

定义的 p 元函数一定是某一随机向量的分布函数，而 $f(\boldsymbol{x})$ 是它相应的分布密度．

例 1.2　二维随机向量 $\boldsymbol{x}=(x_1,x_2)'$ 具有密度

$$f(x_1,x_2)=\begin{cases}e^{-(x_1+x_2)} & x_1>0,\ x_2>0,\\ 0 & \text{其他},\end{cases}$$

那么容易验证 $f(x_1,x_2)$ 满足性质 i) 和 ii)．且此时相应的分布函数为

$$F(x_1,x_2)=\begin{cases}1-e^{-x_1}-e^{-x_2}+e^{-(x_1+x_2)} & x_1>0,x_2>0,\\ 0 & \text{其他}.\end{cases}$$

最重要的多元分布——多元正态分布将在下节详细研究．

（二） 边缘分布、条件分布和统计独立性

既然随机向量 $\boldsymbol{x}$ 的分布函数完全刻划了它的统计特性，那么如何由 $\boldsymbol{x}$ 的分布函数求得它的子向量的分布函数呢？ 为了方便地用符号表达，我们总假定取子向量是由 $\boldsymbol{x}$ 的前 r 个分量组成的（或后 $p-r$ 个分量组成的）．即

$\boldsymbol{x}=(\boldsymbol{x}_1',\ \boldsymbol{x}_2')'$ 其中 $\boldsymbol{x}_1=(x_1,\cdots,\ x_r)'$ 是 r 维向量，$\boldsymbol{x}_2=(x_{r+1},\ \cdots,\ x_p)'$ 是 $p-r$ 维向量，记 $\boldsymbol{x}_1$ 的联合分布函数为 $F_1(\boldsymbol{x}_1)$．于是对任一 r 维实数向量 $\boldsymbol{y}_1=(y_1,\cdots,y_r)'$，事件

$$\{x_1<y_1,\cdots,x_r<y_r\}=\{x_1<y_1,\cdots,x_r<y_r,\\ y_{r+1}<+\infty,\cdots,x_p<+\infty\}.$$

由分布函数定义及概率的连续性可知

$$\begin{aligned}F_1(\boldsymbol{y}_1)&=F_1(y_1,\cdots,y_r)=P(x_1<y_1,\cdots,x_r<y_r)\\&=P(x_1<y_1,\cdots,x_r<y_r,x_{r+1}<+\infty,\cdots,x_p<+\infty)\\&=\lim_{\substack{y_j\to\infty\\ j=r+1,\cdots,p}}F(y_1,\cdots,y_r,y_{r+1},\cdots,y_p)\end{aligned}$$

$$= F(y_1,\cdots,y_r,+\infty,\cdots,+\infty).$$

特别 $\boldsymbol{x}$ 的任一分量 x_i 的边缘分布为

$$F_i(y_i) = F(+\infty,\cdots,+\infty,y_i,+\infty,\cdots,+\infty)$$
$$(i = 1,\cdots,p).$$

对于离散型随机向量 $\boldsymbol{x}$，若它的概率分布为

$$P(x_1 = a_1^{(i_1)},\cdots,x_p = a_p^{(i_p)}) = p(i_1,\cdots,i_p), \qquad (1.4)$$

那么 $\boldsymbol{x}_1 = (x_1,\cdots,x_r)'$ 的边缘概率分布为

$$P_1(x_1 = a_1^{(i_1)},\cdots,x_r = a_r^{(i_r)}) = \sum_{i_j,(j=r+1,\cdots,p)} p(i_1,\cdots,i_p)$$
$$= p_1(i_1,\cdots,i_r).$$

如果 $\boldsymbol{x}$ 是连续型随机向量，它的分布密度为 $f(\boldsymbol{x}) = f(x_1,\cdots,x_p)$，那么 $\boldsymbol{x}_1 = (x_1,\cdots,x_r)'$ 的边缘分布函数为

$$F_1(x_1,\cdots,x_r) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_r}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(t_1,\cdots,t_p)dt_1\cdots dt_p,$$

所以 $\boldsymbol{x}_1$ 的边缘分布密度为

$$f_1(\boldsymbol{x}_1) = f_1(x_1,\cdots,x_r) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\cdots,x_r,$$
$$t_{r+1},\cdots t_p)dt_{r+1}\cdots dt_p. \qquad (1.5)$$

这里要注意：如果 $\boldsymbol{x}$ 有密度，则它的任何子向量的边缘分布都存在，且可由(1.5)来求出相应边缘密度．反之，若 $\boldsymbol{x}$ 的边缘分布都有密度存在，并不能推出 $\boldsymbol{x}$ 有密度．另外，即使 $f(\boldsymbol{x})$ 形式十分简单，由于(1.5)式是多重积分，边缘密度 $f_1(\boldsymbol{x}_1)$ 形式可能很复杂，甚至于没有好的解析表达式．

例 1.3　在例 1.2 中可以计算出 x_1 和 x_2 的边缘分布函数为

$$F_1(x_1) = 1 - e^{-x_1}\ (x_1 > 0),\quad F_2(x_2) = 1 - e^{-x_2}\ (x_2 > 0),$$

它们的边缘分布密度为

$$f_1(x_1) = e^{-x_1}\ (x_1 > 0),\quad f_2(x_2) = e^{-x_2}\ (x_2 > 0).$$

例 1.4　设 $\boldsymbol{x} = (x_1,\cdots,x_{r-1}) \sim PN(n,p_1,\cdots,p_{r-1})$，那么对任一 $s(1 \leqslant s < r)$ 有

i) $(x_1,\cdots,x_{s-1})' \sim PN(n,p_1,\cdots,p_{s-1})$;

ii) $x_i \sim B(n, p_i)\ i = 1, \cdots, r-1$.
具体验证留给读者自行完成。

下面我们复习有关条件分布的概念，条件分布是一个很重要的概念，它是条件概率的推广。

若 A 和 B 是事件，且 $P(B) > 0$，在 B 发生条件下 A 的条件概率定义为

$$P(A \mid B) = \frac{P(AB)}{P(B)} \qquad (P(B) > 0).$$

取 $A = \{a_1 \leqslant x_1 < b_1\}$，$B = \{a_2 \leqslant x_2 < b_2\}$ 其中 x_1, x_2 是随机变量，$a_i \leqslant b_i\ (i = 1, 2)$，且 $P(B) > 0$，于是

$$P(a_1 \leqslant x_1 < b_1 \mid a_2 \leqslant x_2 < b_2) = \frac{P(a_1 \leqslant x_1 < b_1, a_2 \leqslant x_2 < b_2)}{P(a_2 \leqslant x_2 < b_2)}. \tag{1.6}$$

设 $\boldsymbol{x} = (x_1, x_2)'$ 是一个二维随机向量，我们考虑当 x_2 取定 y_2 时的分布，即 $P(x_1 < y_1 \mid x_2 = y_2)$ 称为 x_1 在 $x_2 = y_2$ 下的条件分布。

若 $\boldsymbol{x}$ 是离散型的，它的概率分布为

$$P(x_1 = a_1^{(i_1)},\ x_2 = a_2^{(i_2)}) = p(i_1, i_2).$$

x_2 的边缘概率分布为 $P(x_2 = a_2^{(i_2)}) = p_2(i_2) = \sum_{i_1} p(i_1, i_2)$，那么在 x_2 取定 $a_2^{(i_2)}$ 时 x_1 的条件概率分布为

$$P(x_1 = a_1^{(i_1)} \mid x_2 = a_2^{(i_2)}) = \frac{p(i_1, i_2)}{p_2(i_2)} \qquad (p_2(i_2) > 0).$$

若 $\boldsymbol{x}$ 是连续型的，它的密度为 $f(x_1, x_2)$，而 x_2 的边缘密度为 $f_2(x_2)$，那么由(1.6)式知

$$P(a_1 \leqslant x_1 < b_1 \mid a_2 \leqslant x_2 < b_2) = \frac{\int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2) dx_1 dx_2}{\int_{a_2}^{b_2} f_2(x_2) dx_2}$$

当 (a_1, a_2) 是 $f(x_1, x_2)$ 的连续点以及 a_2 是 $f_2(x_2)$ 的连续点时，令 $b_2 \to a_2$，$b_1 \to a_1$，利用中值定理可知，当 $f_2(a_2) > 0$ 时，则 $f(a_1|a_2) = f(a_1, a_2)/f_2(a_2)$。用 (x_1, x_2) 代 (a_1, a_2)，得到在 x_2 给定时 x_1 的条件密度为

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} \qquad (f(x_2) > 0) \tag{1.7}$$

更一般地，当 $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$ 有密度 $f(\boldsymbol{x}_1, \boldsymbol{x}_2)$，$\boldsymbol{x}_2$ 有边缘密度 $f_2(\boldsymbol{x}_2)$ 时，当 $\boldsymbol{x}$ 是 $f(\boldsymbol{x})$ 的连续点，$\boldsymbol{x}_2$ 是 $f_2(\boldsymbol{x}_2)$ 的连续点时，则当 $\boldsymbol{x}_2$ 给定时 $\boldsymbol{x}_1$ 的条件密度为

$$f(\boldsymbol{x}_1|\boldsymbol{x}_2) = \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f_2(\boldsymbol{x}_2)} \qquad f_2(\boldsymbol{x}_2) > 0 \tag{1.8}$$

若 $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$ 是离散型的，则 $\boldsymbol{x}_2$ 的边缘分布也是离散型的，设 $\boldsymbol{x}$ 和 $\boldsymbol{x}_2$ 的概率分布分别为

$$P(x_1 = a_1^{(i_1)}, \cdots, x_p = a_p^{(i_p)}) = p(i_1, \cdots, i_p)$$

和
$$P(x_{r+1} = a_{r+1}^{(i_{r+1})}, \cdots, x_p = a_p^{(i_p)}) = p_2(i_{r+1}, \cdots, i_p),$$

那么在 $x_{r+1} = a_{r+1}^{(i_{r+1})}, \cdots, x_p = a_p^{(i_p)}$ 条件下 $\boldsymbol{x}_1$ 的条件概率分布为

$$P(x_1 = a_1^{(i_1)}, \cdots, x_r = a_r^{(i_r)} | x_{r+1} = a_{r+1}^{(i_{r+1})}, \cdots, x_p = a_p^{(i_p)})$$
$$= \frac{p(i_1, \cdots, i_p)}{p_2(i_{r+1}, \cdots, i_p)}. \tag{1.9}$$

最后我们讨论随机向量的相互独立性

设 $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$ 是 p 维随机向量，其中子向量 $\boldsymbol{x}_1$ 是 r 维的，$\boldsymbol{x}_2$ 是 $p-r$ 维的，$\boldsymbol{x}$ 的分布函数为 $F(\boldsymbol{x})$，$\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 的边缘分布函数分别为 $F_1(\boldsymbol{x}_1)$ 和 $F_2(\boldsymbol{x}_2)$。

定义 1.3 如果

$$F(\boldsymbol{x}) = F_1(\boldsymbol{x}_1) \cdot F_2(\boldsymbol{x}_2) \tag{1.10}$$

则称 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 是相互独立的。

若 $\boldsymbol{x}$ 是连续型的，$\boldsymbol{x}$ 的分布密度为 $f(\boldsymbol{x})$，$\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 的边缘分布密度为 $f_1(\boldsymbol{x}_1)$ 和 $f_2(\boldsymbol{x}_2)$，那么(1.10)等价于

$$f(\boldsymbol{x}) = f_1(\boldsymbol{x}_1) \cdot f_2(\boldsymbol{x}_2)^{(*)} \tag{1.11}$$

若 $\boldsymbol{x}$ 是离散型的，记 $\boldsymbol{a} = (\boldsymbol{a}_1', \boldsymbol{a}_2')'$，$\boldsymbol{a}_1 = (a_1^{(i_1)}, \cdots, a_r^{(i_r)})'$ $\boldsymbol{a}_2 = (a_{r+1}^{(i_{r+1})}, \cdots, a_p^{(i_p)})'$. 那么 $\boldsymbol{x}$ 的概率分布与 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$ 的边缘概率分布当且仅当有关系式

$$P(\boldsymbol{x} = \boldsymbol{a}) = P_1(\boldsymbol{x}_1 = \boldsymbol{a}_1) \cdot P_2(\boldsymbol{x}_2 = \boldsymbol{a}_2) \tag{1.12}$$

时 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 是相互独立的。

定义 1.4 设 $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2', \cdots, \boldsymbol{x}_m')'$ 是一 p 维随机向量，其中 $\boldsymbol{x}_i$ 是 q_i 维向量，且 $\sum_{i=1}^{m} q_i = p$，设 $\boldsymbol{x}$ 的分布函数为 $F(\boldsymbol{x})$，$\boldsymbol{x}_i$ 的边缘分布函数为 $F_i(\boldsymbol{x}_i)$ $(i = 1, \cdots, m)$. 如果

$$F(\boldsymbol{x}) = \prod_{i=1}^{m} F_i(\boldsymbol{x}_i) \tag{1.13}$$

则称 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m$ 是相互独立的。

当 x_i 都是一维变数时，我们称 $\boldsymbol{x}$ 的诸分量 $x_1, \cdots, x_p$ 是相互独立的。

当 $\boldsymbol{x}$ 是连续型时，(1.13)等价于

$$f(\boldsymbol{x}) = \prod_{i=1}^{m} f_i(\boldsymbol{x}_i) \tag{1.14}$$

其中 $f(\boldsymbol{x})$ 和 $f_i(\boldsymbol{x}_i)$ 分别是 $\boldsymbol{x}$ 的分布密度和 $\boldsymbol{x}_i$ 的边缘密度。

当 $\boldsymbol{x}$ 是离散型时，如同(1.12)可以给出(1.13)的等价条件。

如果将 $\boldsymbol{x}$ 划分为子向量 $\boldsymbol{x}_1, \cdots \boldsymbol{x}_m$，即 $\boldsymbol{x} = (\boldsymbol{x}_1', \cdots, \boldsymbol{x}_m')'$，任意二个子向量 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 相互独立，则称 $\boldsymbol{x}_1, \cdots \boldsymbol{x}_m$ 是两两独立的。

注意如果 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m$ 是相互独立的，它们必定两两独立，反之则未必成立。

定理 1.2 若 p 维随机向量 $\boldsymbol{x} = (\boldsymbol{x}_1', \cdots, \boldsymbol{x}_m')'$ 其中 $\boldsymbol{x}_i$ 的维

(*) 严格说：(1.11)式在 R^p 上几乎处处成立，以后如有类似的地方就不再一一说明。

数为 q_i，$\sum_{i=1}^{m} q_i = p$，则 x_1，$\cdots$，x_m 相互独立的充分必要条件是：

对于 x_i 值域中的任意集合 $B_i(i=1,\cdots,m)$，成立等式

$$P(x_1 \in B_1, \cdots x_m \in B_m) = \prod_{i=1}^{m} P(x_i \in B_i) \qquad (1.15)$$

定理 1.3 若 p 维随机向量 $x=(x_1',\cdots,x_m')'$，其中 x_i 的维数为 q_i，$\sum_{i=1}^{m} q_i = p$，则 $x_1,\cdots x_m$ 相互独立的充分必要条件是：

对于 x_i 的任意 q_i 维可测函数 $g_i(x_i)$ $(i=1,\cdots,m)$，$g_1(x_1),\cdots,g_m(x_m)$ 是相互独立的。

上述二定理的证明可参考[35]。

（三） 多元随机向量的数字特征和特征函数

设 $x=(x_1,\cdots,x_p)'$ 是 p 维随机向量，$Z=(z_{ij})$ 是 $p\times q$ 阶随机矩阵。

定义 1.5 若 p 维随机向量 x 的每个分量 x_i 都有数学期望（均值）Ex_i 存在，则

$$\mu = Ex = (Ex_1,\cdots,Ex_p)' \qquad (1.16)$$

为 x 的数学期望向量（均值向量）。

而随机矩阵 $Z=(z_{ij})_{p\times q}$ 的数学期望矩阵同样可定义为

$$EZ = (Ez_{ij})_{p\times q} \qquad (1.17)$$

定义 1.6 若 p 维随机向量 x 的各分量 x_i 存在二阶矩，其中 x_i 的方差为

$$\mathrm{var}(x_i) = E(x_i - Ex_i)^2 = \sigma_i^2 \triangleq \sigma_{ii} \quad (i=1,\cdots,p).$$

x_i 与 x_j 的协方差为

$$\mathrm{cov}(x_i,x_j) = E(x_i - Ex_i)(x_j - Ex_j) \triangleq \sigma_{ij} \quad (i\neq j; i、j=1,\cdots,p),$$

则称以 σ_{ij} 诸元素组成的 $p\times p$ 阶矩阵

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \cdots & \cdots & \cdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} = (\sigma_{ij}) = E(\boldsymbol{x} - E\boldsymbol{x})(\boldsymbol{x} - E\boldsymbol{x})' \tag{1.18}$$

为 $\boldsymbol{x}$ 的(自)协方差矩阵，简记为 $\text{cov}(\boldsymbol{x}) = \Sigma$.

若 p 维随机向量 $\boldsymbol{x} = (x_1, \cdots, x_p)'$ 和 q 维随机向量 $\boldsymbol{y} = (y_1, \cdots, y_q)'$ 之间诸分量 x_i 和 y_j 之间协方差 $v_{ij} = \text{cov}(x_i, y_j)$ 存在 $(i = 1, \cdots, p;\ j = 1, \cdots, q)$，则称以 v_{ij} 为元素的 $p \times q$ 阶矩阵

$$V = (v_{ij})_{p\times q} = E(\boldsymbol{x} - E\boldsymbol{x})(\boldsymbol{y} - E\boldsymbol{y})'$$

为 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的协方差矩阵，简记为 $\text{cov}(\boldsymbol{x}, \boldsymbol{y}) = V$.

关于求期望和协方差运算有下列定理.

定理 1.4 设 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{v}$ 为 p 维随机向量，Z 是 $p \times q$ 阶随机矩阵，A, B, C, D 是适当的常数矩阵，$\boldsymbol{a}$ 是 p 维常数向量，那么下列等式成立:

1) $E(A\boldsymbol{x} + B\boldsymbol{y}) = AE\boldsymbol{x} + BE\boldsymbol{y}$;

2) $\text{cov}(\boldsymbol{x} + \boldsymbol{a}) = \text{cov}(\boldsymbol{x})$;

3) $\text{cov}(A\boldsymbol{x}, B\boldsymbol{y}) = A\text{cov}(\boldsymbol{x}, \boldsymbol{y})B'$;

4) $E(AZB + C) = A \cdot EZ \cdot B + C$;

5) $E(\text{tr}AZ) = \text{tr}(A \cdot EZ)$;

6) $\text{cov}(A\boldsymbol{x}) = A\text{cov}(\boldsymbol{x})A'$;

7) $\text{cov}(A\boldsymbol{x} + B\boldsymbol{y},\ C\boldsymbol{u} + D\boldsymbol{v}) = A\text{cov}(\boldsymbol{x}, \boldsymbol{u})C' + A\text{cov}(\boldsymbol{x}, \boldsymbol{v})D' + B\text{cov}(\boldsymbol{y}, \boldsymbol{u})C' + B\text{cov}(\boldsymbol{y}, \boldsymbol{v})D'$

证明 在此仅证明 3) 式，其余留给读者证明.

$$\begin{aligned} \text{cov}(A\boldsymbol{x}, B\boldsymbol{y}) &= E[A\boldsymbol{x} - E(A\boldsymbol{x})][B\boldsymbol{y} - E(B\boldsymbol{y})]' \\ &= E\{A(\boldsymbol{x} - E\boldsymbol{x})(\boldsymbol{y} - E\boldsymbol{y}) \cdot B'\} \\ &= AE(\boldsymbol{x} - E\boldsymbol{x})(\boldsymbol{y} - E\boldsymbol{y})'B' \\ &= A\text{cov}(\boldsymbol{x}, \boldsymbol{y})B'. \quad \blacksquare \end{aligned}$$

为了了解 p 维随机向量 $x=(x_1,\cdots,x_p)'$ 各分量 x_i 和 x_j 之间的相关关系，定义 x_i 和 x_j 的相关系数为

$$\rho(x_i,x_j)=\rho_{ij}=\frac{\operatorname{cov}(x_i,x_j)}{[\operatorname{var}(x_i)\operatorname{var}(x_j)]^{\frac{1}{2}}}=\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

$$(\sigma_{ii}>0,\sigma_{jj}>0). \quad (1.19)$$

以 ρ_{ij} 为元素的 $p\times p$ 阶矩阵 $\rho=(\rho_{ij})$ 称为 x 的相关系数矩阵(相关矩阵)。

显然 $\rho_{ii}=1\ (i=1,\cdots,p)$。若记 $D_\sigma=\operatorname{diag}(\sigma_{11},\cdots,\sigma_{pp})$，相关系数矩阵可由下式求得

$$\rho=D_\sigma^{-\frac{1}{2}}\cdot\Sigma\cdot D_\sigma^{-\frac{1}{2}} \quad (1.20)$$

由于 Σ 是非负定矩阵，从而 ρ 是对角元素为 1 的非负定矩阵。

对于随机矩阵 Z，利用 $\operatorname{vec}Z$，即可求得 Z 的诸分量的协方差矩阵和相关系数矩阵。

x_i 和 x_j 的相关系数 ρ_{ij} 具有下列性质(参考[1])：

1) $0\leqslant|\rho_{ij}|\leqslant 1$;

2) 当 x_i,x_j 相互独立时，$\rho_{ij}=0$;

3) $|\rho_{ij}|=1$ 的充分必要条件是存在常数 $a_i\neq 0$ 和 b_i 使得

$$P(x_j=a_ix_i+b_i)=1;$$

4) 对于任意常数 a_i, a_j, b_i, b_j, 设 $y_1=a_ix_i+b_i$, $y_2=a_jx_j+b_j$，那么 y_1 和 y_2 的相关系数 $\rho(y_1,y_2)$ 满足

$$\rho(y_1,y_2)=\begin{cases}\rho_{ij} & a_ia_j>0\\ -\rho_{ij} & a_ia_j<0\\ 0 & a_ia_j=0\end{cases}$$

若 $\rho_{ij}=0$，则称 x_i 和 x_j 是不相关的。

设 $g(x)$ 是 p 维随机向量 x 的可测函数，且 x 的分布函数为 $F(x)$，则 $g(x)$ 的数学期望定义为

$$Eg(x)=\int_{R^p}g(x)dF(x).$$

特殊地，当 $g(\boldsymbol{x})=\prod_{i=1}^{p} x_i^{h_i}$，那么 $Eg(\boldsymbol{x})$ 称为 $\boldsymbol{x}$ 分量的联合矩。若 $\sum_{i=1}^{p} h_i=k$，上述 $Eg(\boldsymbol{x})$ 称为 $\boldsymbol{x}$ 的 k 阶联合矩。

定理 1.5 设 p 维随机向量 $\boldsymbol{x}$ 各分量的二阶矩存在，A 是 $p\times p$ 阶常数对称矩阵，$E\boldsymbol{x}=\boldsymbol{\mu}$，$\operatorname{cov}(\boldsymbol{x})=\Sigma$，那么

1) $E\boldsymbol{x}\boldsymbol{x}'=\Sigma+\boldsymbol{\mu}\cdot\boldsymbol{\mu}'$ (1.21)

2) $E(\boldsymbol{x}'A\boldsymbol{x})=\operatorname{tr}A\Sigma+\boldsymbol{\mu}'A\boldsymbol{\mu}$ (1.22)

证明 1)
$$
\begin{aligned}
E\boldsymbol{x}\boldsymbol{x}' &= E[(\boldsymbol{x}-\boldsymbol{\mu})+\boldsymbol{\mu}][(\boldsymbol{x}-\boldsymbol{\mu})+\boldsymbol{\mu}]' \\
&= E(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})'+E(\boldsymbol{x}-\boldsymbol{\mu}) \\
&\quad \cdot\boldsymbol{\mu}'+\boldsymbol{\mu}\cdot E(\boldsymbol{x}-\boldsymbol{\mu})'+\boldsymbol{\mu}\cdot\boldsymbol{\mu}' \\
&= \Sigma+\boldsymbol{\mu}\cdot\boldsymbol{\mu}'
\end{aligned}
$$

2)
$$
\begin{aligned}
E\boldsymbol{x}'A\boldsymbol{x} &= E(\operatorname{tr}A\boldsymbol{x}\boldsymbol{x}')=\operatorname{tr}A(\Sigma+\boldsymbol{\mu}\boldsymbol{\mu}') \\
&= \operatorname{tr}A\Sigma+\boldsymbol{\mu}'A\boldsymbol{\mu}.
\end{aligned}
$$

定理 1.6 如果 p 维随机向量 $\boldsymbol{x}=(\boldsymbol{x}_1',\ \boldsymbol{x}_2')'$，其 r 维子向量 $\boldsymbol{x}_1$ 和 $p-r$ 维子向量 $\boldsymbol{x}_2$ 相互独立，$g_1(\boldsymbol{x}_1)$ 是 $\boldsymbol{x}_1$ 的可测函数，$g_2(\boldsymbol{x}_2)$ 是 $\boldsymbol{x}_2$ 的可测函数，则有

$$E[g_1(\boldsymbol{x}_1)\cdot g_2(\boldsymbol{x}_2)]=Eg_1(\boldsymbol{x}_1)\cdot Eg_2(\boldsymbol{x}_2) \tag{1.23}$$

证明 设 $\boldsymbol{x}$ 的分布函数为 $F(\boldsymbol{x})$，$\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 的边缘分布函数为 $F_1(\boldsymbol{x}_1)$ 和 $F_2(\boldsymbol{x}_2)$，由 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 相互独立知 $F(\boldsymbol{x})=F_1(\boldsymbol{x}_1)\cdot F_2(\boldsymbol{x}_2)$ 所以

$$
\begin{aligned}
E[g_1(\boldsymbol{x}_1)g_2(\boldsymbol{x}_2)] &= \int_{R^p} g_1(\boldsymbol{x}_1)g_2(\boldsymbol{x}_2)dF(\boldsymbol{x}) \\
&= \int_{R^r} g_1(\boldsymbol{x}_1)dF_1(\boldsymbol{x}_1)\int_{R^{p-r}} g_2(\boldsymbol{x}_2)dF_2(\boldsymbol{x}_2) \\
&= Eg_1(\boldsymbol{x}_1)\cdot Eg_2(\boldsymbol{x}_2).
\end{aligned}
$$

推论 1.1 若随机向量 $\boldsymbol{x}=(x_1,\cdots,x_p)'$ 诸分量相互独立，$g_i(x_i)$ 是 x_i 的可测函数，那么

$$E\left[\prod_{i=1}^{p} g_i(x_i)\right]=\prod_{i=1}^{p} Eg_i(x_i)$$

从一般数理统计中我们已经知道特征函数在理论推导、尤其在极限理论中起了重要作用，这里我们介绍一些多元随机向量的特征函数的定义及其性质，这些性质的证明可参考[2]。

定义 1.7 设 p 维随机向量 $\boldsymbol{x}=(x_1,\cdots,x_p)'$ 具有分布函数 $F(\boldsymbol{x})$，$\boldsymbol{t}=(t_1,\cdots,t_p)'$ 是任一 p 维实数向量，$\boldsymbol{x}$ 的特征函数定义为

$$\varphi_{\boldsymbol{x}}(\boldsymbol{t})=Ee^{i\boldsymbol{t}'\boldsymbol{x}}=Ee^{i\sum_{j=1}^{p}t_jx_j}=\int_{R^p}e^{i\boldsymbol{t}'\boldsymbol{x}}dF(\boldsymbol{x}) \qquad (1.24)$$

多元特征函数 $\varphi_{\boldsymbol{x}}(\boldsymbol{t})$ 具有下列性质(类似于一元情况)

1) $\varphi_{\boldsymbol{x}}(\boldsymbol{t})$ 在 R^p 中一致连续，且 $|\varphi_{\boldsymbol{x}}(\boldsymbol{t})|\leqslant\varphi_{\boldsymbol{x}}(\boldsymbol{0})=1$，$\varphi_{\boldsymbol{x}}(-\boldsymbol{t})=\overline{\varphi_{\boldsymbol{x}}(\boldsymbol{t})}$，其中 $\overline{\varphi_{\boldsymbol{x}}(\boldsymbol{t})}$ 是 $\varphi_{\boldsymbol{x}}(\boldsymbol{t})$ 的共轭函数。

2) 设 $\boldsymbol{A}$ 是 $r\times p$ 阶常数矩阵，$\boldsymbol{b}$ 是 r 维常数向量，那么 $\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$ 的特征函数为

$$\varphi_{\boldsymbol{y}}(\boldsymbol{t})=e^{i\boldsymbol{t}'\boldsymbol{b}}\cdot\varphi_{\boldsymbol{x}}(\boldsymbol{A}'\boldsymbol{t}). \qquad (1.25)$$

3) (边缘分布的特征函数)设 $\boldsymbol{x}=(\boldsymbol{x}_1',\boldsymbol{x}_2')'$，$\boldsymbol{x}_1$ 是 r 维子向量，$\boldsymbol{x}$ 的特征函数为 $\varphi_{\boldsymbol{x}}(\boldsymbol{t})$，$\boldsymbol{x}_1$ 的特征函数为 $\varphi_{\boldsymbol{x}_1}(\boldsymbol{t}^{(1)})$，那末

$$\varphi_{\boldsymbol{x}_1}(\boldsymbol{t}^{(1)})=\varphi_{\boldsymbol{x}}(t_1,\cdots,t_r,0,0,\cdots,0)$$

即在 $\varphi_{\boldsymbol{x}}(\boldsymbol{t})$ 中令 $t_{r+1}=\cdots=t_p=0$。

4) 若 $\boldsymbol{x}$ 的 k 阶联合矩 $E(x_1^{h_1}\cdots x_p^{h_p})$ 存在，$\left(\sum_{i=1}^{p}h_i=k\right)$，则

$$E(x_1^{h_1}\cdots x_p^{h_p})=(-i)^k\left[\frac{\partial^k\varphi_{\boldsymbol{x}}(t_1,\cdots,t_p)}{\partial t_1^{h_1}\cdots\partial t_p^{h_p}}\right]_{t_1=\cdots=t_p=0}. \qquad (1.26)$$

5) 设 $\boldsymbol{x}=(\boldsymbol{x}_1',\boldsymbol{x}_2')'$，且 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 相互独立，$\varphi_{\boldsymbol{x}}(\boldsymbol{t}),\varphi_{\boldsymbol{x}_1}(\boldsymbol{t}_1)$，$\varphi_{\boldsymbol{x}_2}(\boldsymbol{t}_2)$ 分别表示 $\boldsymbol{x},\boldsymbol{x}_1,\boldsymbol{x}_2$ 相应的特征函数，其中 $\boldsymbol{t}=(\boldsymbol{t}_1',\boldsymbol{t}_2')'$ 那么

$$\varphi_{\boldsymbol{x}}(\boldsymbol{t})=\varphi_{\boldsymbol{x}_1}(\boldsymbol{t}_1)\cdot\varphi_{\boldsymbol{x}_2}(\boldsymbol{t}_2), \qquad (1.27)$$

反之，若(1.27)式成立，则 x_1 和 x_2 相互独立。

6）设两个 p 维随机向量 x 和 y 相互独立，则 $z=x+y$ 的特征函数 $\varphi_z(t)$ 满足

$$\varphi_z(t)=\varphi_x(t)\cdot\varphi_y(t),$$

其中 $\varphi_x(t),\varphi_y(t)$ 分别是 x 和 y 的特征函数。

7）设 $x=(x_1,\cdots,x_p)'$ 的特征函数为 $\varphi_x(t)$，其分量 x_j 的特征函数为 $\varphi_{xj}(t_j)$，则 x 的诸分量 $x_1,\cdots,x_p$ 相互独立的充分必要条件为

$$\varphi_x(t)=Ee^{i\sum_{j=1}^{p}t_jx_j}=\prod_{j=1}^{p}Ee^{it_jx_j}=\prod_{j=1}^{p}\varphi x_j(t_j).$$

8）**(逆转公式)** 设 $x=(x_1,\cdots,x_p)'$ 特征函数为 $\varphi_x(t)$，它的分布函数为 $F(x)$ 而 $I_{[a,b)}=\{(x_1,\cdots,x_p):a_k\leqslant x_k<b_k,\ a_k<b_k,\ k=1,\cdots,p\}$ 为 p 维立方体，则

$$P(x\in I_{[a,b)})=\lim_{s_j\to\infty(j=1,\cdots,p)}(2\pi)^{-p}\int_{-s_p}^{s_p}\cdots\int_{-s_1}^{s_1}\prod_{k=1}^{p}$$

$$\cdot\frac{e^{-it_ka_k}-e^{-it_kb_k}}{it_k}\varphi_x(t_1,\cdots,t_p)dt_1\cdots dt_p.$$

这里 a_k,b_k 取法要求 x 在 $I_{[a,b)}$ 的各个面上概率为 0。

9）**(唯一性定理)** 随机向量的分布函数由它的特征函数唯一决定，即两个随机向量的分布函数恒等的充分必要条件是它们的特征函数恒等。

10）**(连续性定理)** 设 $\{F_j(x)\}$是一串分布函数列，$\{\varphi_j(t)\}$是相应的特征函数列，则 $F_j(x)$ 按分布收敛(*)于一个分布函数 $F(x)$ 的充分必要条件是 $\varphi_j(t)$ 收敛到一个在 $t=0$ 处连续的函数 $\varphi(t)$，这时 $F(x)$ 相应的特征函数就是 $\varphi(t)$。

(*) 如果在 $F(x)$的连续点处都有 $\lim_{n\to\infty}F_n(x)=F(x)$，则称 $F_n(x)$ 按分布收敛到 $F(x)$.

例 1.5 设 $\tilde{x}=(x_1,\cdots,x_{r-1})\sim PN(n,p_1,\cdots,p_{r-1})$，求 $\tilde{x}$ 的特征函数。

解
$$\varphi_{\tilde{x}}(\tilde{t})=Ee^{i\sum_{j=1}^{r-1}t_jx_j}$$
$$=\sum_{m_1+\cdots+m_r=n}\frac{n!}{m_1!\cdots m_{r-1}!m_r!}e^{i(t_1m_1+\cdots+t_{r-1}m_{r-1})}p_1^{m_1}\cdots p_r^{m_r}$$
$$=(p_1e^{it_1}+p_2e^{it_2}+\cdots+p_{r-1}e^{it_{r-1}}+p_r)^n.$$

上式当 $t_2=\cdots=t_{r-1}=0$ 时正是二项分布 $B(n,p)$ 的特征函数。

例 1.6 设 $x=(x_1,x_2)'$ 具有例(1.2)中的密度，求它的特征函数和一、二阶矩。

解
$$\varphi_x(t_1,t_2)=\int_0^\infty\int_0^\infty e^{i(t_1x_1+t_2x_2)}\cdot e^{-(x_1+x_2)}dx_1dx_2$$
$$=\frac{1}{(it_1-1)(it_2-1)},$$
$$Ex_1=(-i)\frac{\partial}{\partial t_1}\varphi_x(t_1,t_2)|_{(t_1=0,t_2=0)}$$
$$=-(it_1-1)^{-2}(it_2-1)|_{(t_1=0,t_2=0)}=1.$$

同理有 $Ex_2=1$.
$$Ex_1x_2=(-i)^2\frac{\partial^2}{\partial t_1\partial t_2}\varphi_x(t_1,t_2)|_{(t_1=0,t_2=0)}=1,$$
$$Ex_1^2=(-i)^2\frac{\partial^2}{\partial t_1^2}\varphi_x(t_1,t_2)|_{(t_1=0,t_2=0)}=2.$$

所以
$$Ex=\mu=\begin{pmatrix}1\\1\end{pmatrix},\qquad \Sigma=\mathrm{cov}(x)=\begin{pmatrix}2&1\\1&2\end{pmatrix}-\begin{pmatrix}1\\1\end{pmatrix}(1\quad 1)$$
$$=\begin{pmatrix}1&0\\0&1\end{pmatrix}.$$

在文献[1]，[2]中都提及随机向量及随机向量的函数的变换

及分布，在本段我们将其主要结果归纳一下，并且仅对连续型分布写出其结论。

若 $\boldsymbol{x}=(x_1,\cdots,x_p)'$ 的密度为 $f(\boldsymbol{x})$，$\boldsymbol{y}=(y_1,\cdots,y_m)'$，$(m\leqslant p)$，其中 $y_i=g_i(\boldsymbol{x})$ 是 $\boldsymbol{x}$ 的可测函数，希望求出 m 维随机向量 $\boldsymbol{y}$ 的分布。

设 $F_1(\boldsymbol{y})$ 是 $\boldsymbol{y}$ 的分布函数，则

$$F_1(\boldsymbol{y})=P(g_1(\boldsymbol{x})<y_1\cdots g_m(\boldsymbol{x})<y_m)=P(\boldsymbol{x}\in B)$$

其中 $B=\{(x_1,\cdots,x_p)',\ g_i(x_1,\cdots,x_p)<y_i,\ i=1,\cdots,m\}$，从而

$$F_1(\boldsymbol{y})=\int_{\substack{g_i(\boldsymbol{x})\leqslant y_i\\ i=1,\cdots,m}}\cdots\int f(\boldsymbol{x})d\boldsymbol{x} \tag{1.28}$$

当 $m=1$ 时就得到 p 维随机向量 $\boldsymbol{x}$ 的一维函数的分布。

定理 1.7 设 p 维随机向量 $\boldsymbol{x}=(x_1,\cdots,x_p)'$ 的密度为 $f(\boldsymbol{x})$，而 $\boldsymbol{y}=(y_1,\cdots,y_p)'=g(\boldsymbol{x})=(g_1(\boldsymbol{x}),\cdots,g_p(\boldsymbol{x}))'$ 是从向量 $\boldsymbol{x}$ 到 $\boldsymbol{y}$ 的一对一变换，并设存在唯一的反函数 $\boldsymbol{x}=h(\boldsymbol{y})=(h_1(\boldsymbol{y}),\cdots,h_p(\boldsymbol{y}))'$。它有一阶连续偏导数 $\dfrac{\partial x_i}{\partial y_j}(i,j=1,\cdots,p)$，其变换的 Jacobian 行列式为 $J(\boldsymbol{x}\to\boldsymbol{y})=\left|\dfrac{\partial(\boldsymbol{x})}{\partial(\boldsymbol{y})}\right|_+$，则 $\boldsymbol{y}$ 的分布密度为

$$f_1(\boldsymbol{y})=\begin{cases}f(h(\boldsymbol{y}))\cdot J(\boldsymbol{x}\to\boldsymbol{y}) & \boldsymbol{y}\in g(\boldsymbol{x})\text{ 的值域 }G\\ 0 & \text{其他}\end{cases} \tag{1.29}$$

证明 当 $\boldsymbol{y}\notin g(\boldsymbol{x})$ 的值域 G，显然 $f_1(\boldsymbol{y})=0$，当 $\boldsymbol{y}\in G$ 时由(1.28)知，$\boldsymbol{y}$ 的分布函数为

$$F_1(\boldsymbol{y})=\int_{\substack{g_i(\boldsymbol{x})\leqslant y_i\\ i=1,\cdots,p}}\cdots\int f(\boldsymbol{x})d\boldsymbol{x}$$

利用重积分的变量替换公式知

$$F_1(\boldsymbol{y})=\int_{-\infty}^{y_p}\cdots\int_{-\infty}^{y_1}f(h(\boldsymbol{y}))J(\boldsymbol{x}\to\boldsymbol{y})d\boldsymbol{y}$$

由密度函数定义知 $\boldsymbol{y}$ 的密度为

$$f_1(\boldsymbol{y})=f(h(\boldsymbol{y}))\cdot J(\boldsymbol{x}\to\boldsymbol{y})\quad \boldsymbol{y}\in G. \quad \blacksquare$$

注意,i) 若 $\boldsymbol{y}$ 的维数 m 低于 $\boldsymbol{x}$ 的维数 p,则欲求 $\boldsymbol{y}$ 的密度可先作变换 $\tilde{\boldsymbol{y}}=(\boldsymbol{y}',x_{m+1},\cdots x_p)'$, 即 $\tilde{\boldsymbol{y}}$ 的前 m 个分量为 $\boldsymbol{y}$,后 $p-m$ 个分量为 $x_{m+1},\cdots,x_p$. 如定理 1.7 求出 $\tilde{\boldsymbol{y}}$ 的分布密度,然后再求 $\tilde{\boldsymbol{y}}$ 的子向量 $\boldsymbol{y}$ 的边缘分布密度.

ii) 若 $\boldsymbol{y}=g(\boldsymbol{x})$ 的反函数不唯一,而有 k 组逆变换.

$$\boldsymbol{x}^{(i)}=h^{(i)}(\boldsymbol{y}),\quad i=1,\cdots,k$$

此时 $\boldsymbol{y}$ 取每组值 $(y_1,\cdots,y_p)'$ 将对应于 k 组 $\boldsymbol{x}^{(i)}=(x_1^{(i)},\cdots,x_p^{(i)})$, 因而 $\boldsymbol{y}$ 取 G 中某区域的概率为 $\boldsymbol{x}$ 取值于 k 个相应集合概率之和,于是每组 $\boldsymbol{y}$ 与 $\boldsymbol{x}^{(i)}$ (它们是一一对应的)可用定理 1.7 求出它的密度,然后相加而得到 $\boldsymbol{y}$ 的密度,即

$$f_1(\boldsymbol{y})=\begin{cases}\sum_{i=1}^{k}f(h^{(i)}(\boldsymbol{y}))J_i(\boldsymbol{x}\to\boldsymbol{y}) & \boldsymbol{y}\in g(\boldsymbol{x})\text{ 的值域}G\\ 0 & \text{其他}\end{cases}$$

其中 $J_i(\boldsymbol{x}\to\boldsymbol{y})$ 是第 i 组逆变换的 Jacobian 行列式.

§2.2 多元正态分布

首先回顾一下一元正态分布, 设 $x\sim N(\mu,\sigma^2)$ $(-\infty<\mu<+\infty,\sigma^2>0)$ 即指随机变数 x 具有密度

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{\frac{-1}{2\sigma^2}(x-\mu)^2\right\} \tag{2.1}$$

它的特征函数为

$$\varphi(t)=\exp\left\{it\mu-\frac{1}{2}\sigma^2t^2\right\} \tag{2.2}$$

如果令 $y=\dfrac{x-\mu}{\sigma}$, 那么 $y\sim N(0,1)$ (称 y 服从标准正态分布)此时则有 $x=\mu+\sigma y$ 的关系式成立.

作为一元情况的推广,我们给出多元正态分布的定义.

定义 2.1（p 维正态分布的古典定义）如果 p 维随机向量 $\boldsymbol{x}=(x_1,\cdots,x_p)'$ 具有下列分布密度

$$f(\boldsymbol{x})=(2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \tag{2.3}$$

其中 $\boldsymbol{\mu}$ 是 p 维常数向量，$\boldsymbol{\Sigma}$ 是 p 阶正定矩阵．则称 $\boldsymbol{x}$ 服从 p 元正态分布，记为 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\Sigma)$．

下面验证 $f(\boldsymbol{x})$ 确是一分布密度．

由于 $\Sigma>0$，$|\Sigma|>0$，所以 $f(\boldsymbol{x})>0$，只须证明

$$I=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(\boldsymbol{x})d\boldsymbol{x}=1.$$

作变换 $\boldsymbol{x}=\Sigma^{\frac{1}{2}}\boldsymbol{y}+\boldsymbol{\mu}$，其中 $\boldsymbol{y}=(y_1,\cdots,y_p)'$，由第一章推论 4.1 有 $J(\boldsymbol{x}\to\boldsymbol{y})=|\Sigma^{\frac{1}{2}}|_+=|\Sigma|^{\frac{1}{2}}$．此时 $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=\boldsymbol{y}'\cdot\boldsymbol{y}$，得到

$$I=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(2\pi)^{-p/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{p}y_i^2\right)dy_1\cdots dy_p$$

$$=\prod_{i=1}^{p}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}y_i^2\right\}dy_i=1$$

即 $f(\boldsymbol{x})$ 确实是一个 p 维分布密度．

从上面推导中看到 $\boldsymbol{y}=(y_1,\cdots,y_p)'$ 有密度 $(2\pi)^{-\frac{p}{2}}\exp\cdot\left\{-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y}\right\}$ 即 $\boldsymbol{y}\sim N_p(\boldsymbol{0},I_p)$．它的各个分量服从 $N(0,1)$ 且相互独立，我们称 $\boldsymbol{y}$ 是 p 元标准正态向量，由此可知

$$E\boldsymbol{y}=0,\quad \mathrm{cov}(\boldsymbol{y})=I_p.$$

对于随机向量 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\Sigma)$，它可写成 $\boldsymbol{x}=\Sigma^{\frac{1}{2}}\boldsymbol{y}+\boldsymbol{\mu}$，其中 $\boldsymbol{y}\sim N_p(0,I_p)$，于是

$$E\boldsymbol{x}=\boldsymbol{\mu},\quad \mathrm{cov}(\boldsymbol{x})=\Sigma^{\frac{1}{2}}\cdot\mathrm{cov}(\boldsymbol{y})\cdot\Sigma^{\frac{1}{2}\prime}=\Sigma.$$

由此可见 p 维正态分布完全由它的期望向量 $\boldsymbol{\mu}$ 和协方差矩阵 Σ 决定．

下面求 $\boldsymbol{x}$ 的特征函数。

设 $\boldsymbol{t}=(t_1,\cdots,t_p)'$ 是任意 p 维实向量，$\boldsymbol{x}$ 特征函数为

$$\varphi_{\boldsymbol{x}}(\boldsymbol{t})=E\exp(i\boldsymbol{t}'\boldsymbol{x})=E\exp[i\boldsymbol{t}'(\Sigma^{\frac{1}{2}}\boldsymbol{y}+\boldsymbol{\mu})]$$

令 $\boldsymbol{a}=\Sigma^{\frac{1}{2}}\cdot\boldsymbol{t}$，由 2.21知道

$$\varphi_{\boldsymbol{x}}(\boldsymbol{t})=\exp(i\boldsymbol{t}'\boldsymbol{\mu})\cdot E\exp(i\boldsymbol{a}'\boldsymbol{y})=\exp(i\boldsymbol{t}'\boldsymbol{\mu})\cdot\prod_{\nu=1}^{p}E\exp(ia_\nu y_\nu)$$

$$=\exp(i\boldsymbol{t}'\boldsymbol{\mu})\cdot\exp\left(-\frac{1}{2}\sum_{\nu=1}^{p}a_\nu^2\right)$$

$$=\exp\left[i\boldsymbol{t}'\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{a}'\boldsymbol{a}\right]=\exp\left[i\boldsymbol{t}'\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right].$$

由特征性质的唯一性，我们也能如下定义 p 元正态分布。

定义 2.2 如果 p 维随机向量 $\boldsymbol{x}$ 有特征函数

$$\varphi_{\boldsymbol{x}}(\boldsymbol{t})=\exp\left\{i\boldsymbol{t}'\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right\}$$

其中 $\boldsymbol{\mu}$ 是 p 维常数向量，Σ 是 p 阶非负定矩阵，则称 $\boldsymbol{x}$ 服从 p 元正态分布 $N_p(\boldsymbol{\mu},\Sigma)$。

注意到在定义 2.2 中只要求 Σ 非负定，显然当 Σ 正定时，它与定义 2.1 等价。而当 $|\Sigma|=0$ 时，设 $\mathrm{rk}(\Sigma)=r<p$，可以证明 $\boldsymbol{x}$ 的分布概率为 1 集中于 R^p 的一个 r 维子空间上，从而无法给出 (2.3)的密度形式，我们称之为退化(或奇异)的正态分布。今后在理论探讨中经常使用定义 2.2，因为它包含了非退化和退化的两种情况。

下面定理指出了多元正态分布与一元正态分布之间的密切联系。

定理 2.1 p 维随机向量 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\Sigma)$ 的充分必要条件是对于任意 p 维常数向量 $\boldsymbol{a}$，线性组合 $y=\boldsymbol{a}'\boldsymbol{x}$ 服从一元正态分布 $N(\boldsymbol{a}'\boldsymbol{\mu},\boldsymbol{a}'\Sigma\boldsymbol{a})$。

证明 由于 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\Sigma)$，它的特征函数为

$$\exp\left\{it'\mu - \frac{1}{2}t'\Sigma t\right\}$$

从而 $y = a'x$ 的特征函数为

$$\varphi_y(\tau) = E(\exp i\tau y) = E\exp[i(\tau a)'x]$$
$$= \exp\left[i\tau(a'\mu) - \frac{1}{2}\tau^2(a'\Sigma a)\right]$$

与(2.2)比较，它正是 $N(a'\mu, a'\Sigma a)$ 的特征函数，由特征函数唯一性知 $y \sim N(a'\mu, a'\Sigma a)$。当 $a'\Sigma a = 0$ 时，有 $P(y = a'\mu) = 1$，则 y 可看成是服从退化正态分布 $N(a'\mu, 0)$。

反之，若对任意 p 维常数向量 $a, y = a'x \sim N(a'\mu, a'\Sigma a)$ 于是 y 的特征函数为 $\varphi_y(\tau) = \exp\left[i\tau a'\mu - \frac{1}{2}\tau^2 a'\Sigma a\right]$

它也可看成

$$E\exp[i(\tau a)'x] = \exp\left[i(\tau a)'\mu - \frac{1}{2}(\tau a)'\Sigma(\tau a)\right].$$

令 $\tau = 1$，由于 a 是任意 p 维向量，$E\exp(ia'x)$ 正是 x 的特征函数

$$\varphi_x(a) = \exp\left[ia'\mu - \frac{1}{2}a'\Sigma a\right]$$

由定义 1.2 知 $x \sim N_p(\mu, \Sigma)$。

由定理 2.1 我们又可给出 p 元正态分布的另一等价定义。

定义 2.3 如果 p 维随机向量 x 的任一线性组合 $a'x$ 服从一元正态分布 $N(a'\mu, a'\Sigma a)$（包括退化的正态分布），则我们称 x 服从 p 元正态分布 $N_p(\mu, \Sigma)$。

定义 2.3 的优点在于把一个多元正态分布的问题化为等价的一元正态分布问题进行讨论，从而可推出多元正态分布的许多性质与一元时相同。

下面将给出多元正态分布在线性变换下的不变性，并讨论它的子向量的边缘分布以及相互的独立性，最后给出独立性和正态

性之间的关系。

定理 2.2 设 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$. 那么对任一 $r \times p$ 阶常数矩阵 D 及任一 r 维常数向量 $\boldsymbol{a}$ 有

$$D\boldsymbol{x} + \boldsymbol{a} \sim N_r(D\boldsymbol{\mu} + \boldsymbol{a}, D'\Sigma D)$$

证明 设 $\boldsymbol{t}$ 是任一 r 维向量，令 $\boldsymbol{b} = D'\boldsymbol{t}$，那么 $D\boldsymbol{x} + \boldsymbol{a}$ 的特征函数是

$$\begin{aligned} E[\exp i\boldsymbol{t}'(D\boldsymbol{x} + \boldsymbol{a})] &= \exp(i\boldsymbol{t}'\boldsymbol{a})E[\exp i\boldsymbol{b}'\boldsymbol{x}] \\ &= \exp(i\boldsymbol{t}'\boldsymbol{a}) \cdot \exp\left(i\boldsymbol{b}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{b}'\Sigma\boldsymbol{b}\right) \\ &= \exp\left[i\boldsymbol{t}'(D\boldsymbol{\mu} + \boldsymbol{a}) - \frac{1}{2}\boldsymbol{t}'(D\Sigma D')\boldsymbol{t}\right] \end{aligned}$$

从而 $D\boldsymbol{x} + \boldsymbol{a} \sim N_r(D\boldsymbol{\mu} + \boldsymbol{a}, D\Sigma D')$.

推论 2.1 令 $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')' \sim N_p(\boldsymbol{\mu}, \Sigma)$，其中

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$\boldsymbol{\mu}_1$ 与 $\boldsymbol{x}_1$ 都是 r 维子向量，Σ_{11} 是 $r \times r$ 阶，那么

$$\boldsymbol{x}_1 \sim N_{r_1}(\boldsymbol{\mu}_1, \Sigma_{11}).$$

证明 只须在定理 2.2 中取 $D = (I_r, 0)_{r \times p}$ 即可。 ▮

由于 $\boldsymbol{x}_1$ 的选取是任意的，且 x_i 位置允许交换，从而有下列定理。

定理 2.3 设 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$，则 $\boldsymbol{x}$ 的任一子向量也是正态分布的，它的期望向量等于 $\boldsymbol{\mu}$ 的相应子向量，而协方差矩阵等于 Σ 相应的子矩阵。

若记 $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)'$，$\Sigma = (\sigma_{ij})_{p \times p}$，则 $\boldsymbol{x}$ 的第 i 个分量 $x_i \sim N(\mu_i, \sigma_{ii})$.

注意到定理2.3的逆定理并不成立，即尽管每个分量的边缘分布是正态分布，但 $\boldsymbol{x} = (x_1, \cdots, x_p)'$ 的联合分布并不一定是正态分布的。

例如，二维随机向量 $\boldsymbol{x} = (x_1, x_2)'$ 有密度

$$f(x_1,x_2)=\alpha[2\pi\sqrt{1-\rho^2}\,]^{-1}\exp\left\{\frac{-1}{2(1-\rho^2)}(x_1^2-2\rho x_1x_2+x_2^2)\right\}$$
$$+(1-\alpha)[2\pi\sqrt{1-\theta^2}\,]^{-1}$$
$$\cdot\exp\left\{\frac{-1}{2(1-\theta^2)}(x_1^2-2\theta x_1x_2+x_2^2)\right\} \tag{2.4}$$

其中 $0<\alpha<1$，$0<|\rho|<1$，$0<|\theta|<1$，$\rho\neq\theta$。容易验证 x_1，x_2 的边缘分布是正态的，但(2.4)却不是正态分布的密度。

定理 2.4 p 维随机向量 $x\sim N_p(\mu,\Sigma)$（Σ 是 p 阶非负定阵，且 $\text{rk}(\Sigma)=r$）的充分必要条件是存在秩为 r 的 $r\times p$ 阶矩阵 B，使得 $BB'=\Sigma$ 且有 $x=By+\mu$，其中 $y\sim N_r(0,I_r)$。

证明　充分性由定理 2.2 立即可得。

现在设 $x\sim N_p(\mu,\Sigma)$，且 $\text{rk}(\Sigma)=r<p$，由于 Σ 非负定，由第一章定理 3.1 知存在正交阵 P 使得

$$P\Sigma P'=\begin{pmatrix}\Lambda & 0\\ 0 & 0\end{pmatrix}$$

其中 $\Lambda=\text{diag}(\lambda_1,\cdots,\lambda_r)$ 是由 Σ 的非零特征根 $\lambda_i>0$ 构成的 r 阶对角阵。令

$$D=\begin{pmatrix}\Lambda^{-\frac{1}{2}} & 0\\ 0 & I_{p-r}\end{pmatrix}$$

显然 D 非奇异，于是 $F=DP$ 也非奇异且成立

$$F\Sigma F'=\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}$$

令 $y=(y_1',y_2')'=F(x-\mu)$，其中 y_1 是 r 维向量，y_2 为 $p-r$ 维向量，由定理 2.2 知 y 是正态分布，且 $Ey=0$，$\text{cov}(y_1)=I_r$，$\text{cov}(y_2)=0$，从而 $P(y_2=0)=1$ 成立而 $y_1\sim N_r(0,I_r)$。记 $F^{-1}=(B,G)$，其中 B 是 $p\times r$ 阶矩阵，于是

$$x-\mu=F^{-1}y=(B,G)\begin{pmatrix}y_1\\ y_2\end{pmatrix}=By_1$$

概率为 1 成立，即 $\boldsymbol{x} = B\boldsymbol{y}_1 + \boldsymbol{\mu}$，此外

$$\Sigma = F^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}(F^{-1})' = (B, G)\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} B' \\ G' \end{pmatrix} = B \cdot B'.$$

从而定理得证。 ∎

定理 2.4 一方面刻划了奇异正态分布与非奇异正态分布之间的联系，另一方面也给出了 p 元正态分布的另一个等价定义。

定义 2.4 若 p 维随机向量 $\boldsymbol{x}$ 可以表示为

$$\boldsymbol{x} = B\boldsymbol{y} + \boldsymbol{\mu}$$

其中 $\boldsymbol{y}$ 是 r 维随机向量且 $\boldsymbol{y} \sim N_r(\boldsymbol{0}, I_r)$。$B$ 是秩为 r 的 $p \times r$ 阶矩阵，$\boldsymbol{\mu}$ 是 p 维常数向量，令 $\Sigma = BB'$，则称 $\boldsymbol{x}$ 服从 p 元正态分布 $N_p(\boldsymbol{\mu}, \Sigma)$。称 r 是这个分布的秩。

定理 2.5 设 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$，并将 $\boldsymbol{x}$ 划分成 $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$ 对 $\boldsymbol{\mu}$ 和 Σ 作相应的划分

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

则 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 相互独立的充分必要条件是 $\Sigma_{12} = 0$。

证明　由于相互独立保证协方差为 0，必要性是显然的。

若 $\Sigma_{12} = 0$，设 $\boldsymbol{t} = (\boldsymbol{t}_1', \boldsymbol{t}_2')'$ 为任一 p 维向量，其中 $\boldsymbol{t}_1$ 是 r 维的。于是

$$\begin{aligned}
\varphi_{\boldsymbol{x}}(\boldsymbol{t}) &= E\exp(i\boldsymbol{t}'\boldsymbol{x}) \\
&= \exp\left\{i(\boldsymbol{t}_1', \boldsymbol{t}_2')\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} - \frac{1}{2}(\boldsymbol{t}_1', \boldsymbol{t}_2')\begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{t}_1 \\ \boldsymbol{t}_2 \end{pmatrix}\right\} \\
&= \exp\left\{i\boldsymbol{t}_1'\boldsymbol{\mu}_1 - \frac{1}{2}\boldsymbol{t}_1'\Sigma_{11}\boldsymbol{t}_1\right\} \\
&\quad \cdot \exp\left\{i\boldsymbol{t}_2'\boldsymbol{\mu}_2 - \frac{1}{2}\boldsymbol{t}_2'\Sigma_{22}\boldsymbol{t}_2\right\} \\
&= E[\exp(i\boldsymbol{t}_1'\boldsymbol{x}_1)] \cdot E[\exp(i\boldsymbol{t}_2'\boldsymbol{x}_2)] \\
&= \varphi_{\boldsymbol{x}_1}(\boldsymbol{t}_1) \cdot \varphi_{\boldsymbol{x}_2}(\boldsymbol{t}_2).
\end{aligned}$$

由特征函数性质 5）知道 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 相互独立。

注意，通常两个随机向量互不相关并不能保证它们的相互独立性，例如在密度(2.4)中令 $\alpha=\frac{1}{2}$，$\rho=-\theta$，可验证 $\operatorname{cov}(x_1, x_2)=0$，但它们并不独立。

推论 2.2 设 $\boldsymbol{x}\sim N_p(\boldsymbol{0}, I_p)$. $\boldsymbol{y}=A\boldsymbol{x}+\boldsymbol{\mu}$，$\boldsymbol{z}=B\boldsymbol{x}+\boldsymbol{\nu}$ 其中 A 是 $n\times p$ 阶矩阵，B 是 $m\times p$ 阶矩阵，$\boldsymbol{\mu}$ 是 n 维向量，$\boldsymbol{\nu}$ 是 m 维向量，那么 $\boldsymbol{y}$ 和 $\boldsymbol{z}$ 相互独立的充分必要条件是 $AB'=0$。

证明　令

$$\boldsymbol{u}=\begin{pmatrix}\boldsymbol{y}\\ \boldsymbol{z}\end{pmatrix}=\begin{pmatrix}A\\ B\end{pmatrix}\boldsymbol{x}+\begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\nu}\end{pmatrix}$$

由定理 2.2 知

$$\boldsymbol{u}\sim N_{n+m}\left(\begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\nu}\end{pmatrix}\begin{pmatrix}AA' & AB'\\ BA' & BB'\end{pmatrix}\right).$$

由定理 2.5 知道 $\boldsymbol{y}$ 与 $\boldsymbol{z}$ 相互独立的充分必要条件正是 $AB'=0$。　▮

推论 2.3 设 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu}, \Sigma)$，$\boldsymbol{y}=A\boldsymbol{x}+\boldsymbol{e}$，$\boldsymbol{z}=B\boldsymbol{x}+\boldsymbol{d}$ 则 $\boldsymbol{y}$ 和 $\boldsymbol{z}$ 相互独立的充分必要条件是 $A\Sigma B'=0$。

证明　按定义 (2.4) 知存在 r 维随机向量 $\boldsymbol{w}\sim N_r(\boldsymbol{0}, I_r)$，$(r=\operatorname{rk}(\Sigma))$ 使得 $\boldsymbol{x}=D\boldsymbol{w}+\boldsymbol{\mu}$ 且 $DD'=\Sigma$。
从而有

$$\boldsymbol{y}=AD\boldsymbol{w}+(A\boldsymbol{\mu}+\boldsymbol{e}),\quad \boldsymbol{z}=BD\boldsymbol{w}+(B\boldsymbol{\mu}+\boldsymbol{d})$$

由推论 2.2 知道 $\boldsymbol{y},\boldsymbol{z}$ 相互独立的充分必要条件是

$$(AD)(BD)'=ADD'B'=A\Sigma B'=0.\qquad ▮$$

推论 2.4 设 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\Sigma)$，且它们的分块形式为

$$\boldsymbol{x}=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\\ \vdots\\ \boldsymbol{x}_k\end{pmatrix}\qquad \boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\\ \vdots\\ \boldsymbol{\mu}_k\end{pmatrix}\qquad \Sigma=\begin{pmatrix}\Sigma_{11} & \cdots & \Sigma_{1k}\\ \vdots & & \\ \Sigma_{k1} & \cdots & \Sigma_{kk}\end{pmatrix}$$

则 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k$ 相互独立的充分必要条件是 Σ 的非对角块 $\Sigma_{ij}=0$

$(i \neq j, i、j = 1,2,\cdots,k)$. 即 $\Sigma = \mathrm{diag}(\Sigma_{11},\cdots,\Sigma_{kk})$.

关于多元正态分布的可加性有如下定理:

定理 2.6 设 $\boldsymbol{x}_r (r = 1, \cdots, k)$ 相互独立且 $\boldsymbol{x}_r \sim N_p(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$, 那么对于已知的 $m \times p$ 阶矩阵 $A_r (r = 1,\cdots,k)$, 有

$$\boldsymbol{y} = \sum_{r=1}^{k} A_r \boldsymbol{x}_r \sim N_m\left(\sum_{r=1}^{k} A_r \boldsymbol{\mu}_r, \sum_{r=1}^{k} A_r \Sigma_r A_r'\right)$$

证明 考虑特征函数

$$\begin{aligned}\varphi_y(\boldsymbol{t}) &= E\exp\left[i\boldsymbol{t}'\left(\sum_{r=1}^{k} A_r \boldsymbol{x}_r\right)\right] \\ &= \prod_{r=1}^{k} E(\exp i\boldsymbol{t}' A_r \boldsymbol{x}_r) \\ &= \prod_{r=1}^{k} \exp\left(i\boldsymbol{t}' A_r \boldsymbol{\mu}_r - \frac{1}{2} \boldsymbol{t}'(A_r \Sigma A_r')\boldsymbol{t}\right) \\ &= \exp\left\{i\boldsymbol{t}' \sum_{r=1}^{k} A_r \boldsymbol{\mu}_r - \frac{1}{2} \boldsymbol{t}'\left(\sum_{r=1}^{k} A_r \Sigma A_r'\right)\boldsymbol{t}\right\}\end{aligned}$$

于是由特征函数唯一性即可推出结论. ▮

推论 2.5 如果 $\boldsymbol{x}_r (r = 1,\cdots,k)$ 相互独立且服从同一分布 $N_p(\boldsymbol{\mu},\Sigma)$, 那么样本均值

$$\bar{\boldsymbol{x}} = \frac{1}{k} \sum_{r=1}^{k} \boldsymbol{x}_r \sim N_p\left(\boldsymbol{\mu}, \frac{1}{k} \Sigma\right).$$

证明 只须在定理 2.6 中取 $m = p$, $A_r = \frac{1}{k} I_p \ (r=1,\cdots,k)$ 即可.

关于随机向量的正态性和独立性有一定的关系，下面两个定理给出了这方面的一些结果.

定理 2.7 设 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 是两个相互独立的 p 维随机向量，则 $\boldsymbol{x},\boldsymbol{y}$ 服从正态分布的充分必要条件是 $\boldsymbol{x}+\boldsymbol{y}$ 服从正态分布.

证明 必要性由定理 2.6 中令 $k = 2$, $A_1 = A_2 = I_p$ 立即可

得。

对于充分性当 $p=1$ 时 Cramer 曾给出了证明，其证明涉及其他知识内容，我们不再证明，我们将利用 $p=1$ 的结论和定义 2.3 证明 $p>1$ 的结果。设 $\boldsymbol{x}$，$\boldsymbol{y}$ 独立且 $\boldsymbol{x}+\boldsymbol{y}$ 服从正态分布，令 $u=\boldsymbol{a}'\boldsymbol{x}$，$v=\boldsymbol{a}'\boldsymbol{y}$（其中 $\boldsymbol{a}$ 是任意非零向量)，则由定理 2.1 知道 $u+v$ 服从一维正态分布，且由定理 1.3 知道 u 和 v 仍保持独立，由 $p=1$ 的结论知道 u，v 服从正态分布，由 $\boldsymbol{a}$ 的任意性，按定义 2.3 知 $\boldsymbol{x},\boldsymbol{y}$ 服从正态分布。 ▮

定理 2.8 设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m$ 是相互独立的 p 维随机向量组，记

$$\boldsymbol{y}=\sum_{i=1}^{m}a_i\boldsymbol{x}_i,\quad \boldsymbol{z}=\sum_{i=1}^{m}b_i\boldsymbol{x}_i,$$

其中 $a_i,b_i(i=1,\cdots,m)$ 是两组实数，则有

1）若 $\boldsymbol{x}_i(i=1,\cdots,m)$ 服从相同的正态分布，且 $\sum_{i=1}^{m}a_ib_i=0$，则 $\boldsymbol{y}$，$\boldsymbol{z}$ 相互独立。

2）若 $\boldsymbol{y},\boldsymbol{z}$ 相互独立，则每个使得 $a_ib_i\neq 0$ 的相应的 $\boldsymbol{x}_i$ 服从 p 维正态分布。

关于定理的 1) 的证明是容易的，2)的证明可参考[15]。

最后我们给出一个二维正态分布的例子。

例 2.1 设 $p=2$，

$$\boldsymbol{x}=\begin{pmatrix}x_1\\x_2\end{pmatrix}\sim N_2\left(\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix},\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$$

1）当 $|\rho|<1$ 时，

$$\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2 & \sigma_2\end{pmatrix}\text{ 满足}$$

$|\boldsymbol{\Sigma}|=(1-\rho^2)\sigma_1^2\sigma_2^2$，$\Sigma>0$，且

$$\boldsymbol{\Sigma}^{-1}=\frac{1}{1-\rho^2}\begin{pmatrix}\sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1}\\-\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2}\end{pmatrix}$$

$\boldsymbol{x}$ 的密度为

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

而 x_1 的边缘分布为 $N(\mu_1,\sigma_1^2)$，x_2 边缘分布为 $N(\mu_2,\sigma_2^2)$，ρ 是 x_1 和 x_2 的相关系数，显然当且仅当 $\rho=0$ 时

$$f(x_1, x_2)=f_1(x_1)f_2(x_2)$$

(其中 $f_1(x_1)$ 和 $f_2(x_2)$ 正是 x_1,x_2 的边缘分布密度)即 x_1，x_2 相互独立。

2) 当 $|\rho|=1$ 时，

此时 $|\Sigma|=0$，可推出 x_1 和 x_2 之间以概率 1 存在着线性关系。这是由于 Σ 非满秩，存在非零向量 $\boldsymbol{t}=(t_1,t_2)'$ 使得 $\Sigma\boldsymbol{t}=0$，由 $\boldsymbol{t}'\Sigma\boldsymbol{t}=0$ 推得 $E[\boldsymbol{t}'(\boldsymbol{x}-\boldsymbol{\mu})]^2=0$。另外又因为 $E[\boldsymbol{t}'(\boldsymbol{x}-\boldsymbol{\mu})]=0$，所以

$\boldsymbol{t}'(\boldsymbol{x}-\boldsymbol{\mu})=t_1(x_1-\mu_1)+t_2(x_2-\mu_2)=0$ 以概率为 1 成立。

反之，若 x_1,x_2 以概率 1 存在线性关系，不妨设 $x_2=ax_1+b(a\neq 0)$，那么

$$\rho=\frac{\operatorname{cov}(x_1, x_2)}{[(\operatorname{var}(x_1)\operatorname{var}(x_2)]^{1/2}}=\frac{a\operatorname{var}(x_1)}{|a|\operatorname{var}(x_1)}=\begin{cases}1 & a>0\\ -1 & a<0.\end{cases}$$

即 $|\rho|=1$。

当 $\rho>1$ 时我们称 x_1 和 x_2 存在正相关，$\rho<0$ 时称 x_1 和 x_2 存在负相关。

为了对多维正态密度的几何表示有更清楚的理解，我们把具有等密度的点的轨迹称为等高线(面)，显然当 $p=2$ 时正是

$$\frac{(x_1-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}$$

$$+\frac{(x_2-\mu_2)^2}{\sigma_2^2}=c^2 \qquad (c\geqslant 0)$$

它是中心在 $(\mu_1,\mu_2)'$ 的一个椭圆。一般 p 维正态等高面为

$$(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=c^2 \quad (c\geqslant 0)$$

它是中心在 $\boldsymbol{\mu}$ 的一个 p 维椭球面。从解析几何和矩阵理论知道由于 $\Sigma>0$，存在正交矩阵 $P=(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_p)$，使得 $\Sigma=P\Lambda P'$。其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_p)$，是由 Σ 的特征根 λ_i 构成的对角阵，$\boldsymbol{e}_i$ 是相应的特征向量。从而 $\Sigma^{-1}=P\Lambda^{-1}P'$，代入等高面方程得到

$$[P'(\boldsymbol{x}-\boldsymbol{\mu})]'\Lambda^{-1}[P'(\boldsymbol{x}-\boldsymbol{\mu})]=c^2,$$

即

$$\sum_{i=1}^{p}\frac{(\boldsymbol{e}_i'\boldsymbol{x}-\boldsymbol{e}_i'\boldsymbol{\mu})^2}{\lambda_i}=c^2$$

若令 $y_i=\boldsymbol{e}_i'\boldsymbol{x}$，$\nu_i=\boldsymbol{e}_i'\boldsymbol{\mu}$，此即标准椭球面的方程，其半轴轴长度分别为$\sqrt{\lambda_i}\,c$，而轴的方向正是 $\boldsymbol{e}_i$，椭球的中心坐标为 $(\boldsymbol{e}_1'\boldsymbol{\mu}\cdots,\boldsymbol{e}_p'\boldsymbol{\mu})'$。

§2.3 正态向量的条件分布和相关性

在第一节中已介绍了随机向量的相关系数 ρ_{ij} 及相关矩阵的概念。ρ_{ij} 表示 $\boldsymbol{x}$ 的分量 x_i 和 x_j 的相关程度。本节将进一步介绍随机向量的其他相关性度量，然后再对正态分布进行讨论，由于我们仅对协方差矩阵感兴趣，不妨假设随机向量的期望向量总为零向量。

（一）偏相关和复相关

相关系数仅反映了分量 x_i 和 x_j 之间的相关程度，而在实际问题中常会遇到其他一些相关问题，例如某一分量与其他分量的全体之间的相关关系，或者在消去若干分量的影响后，其余分量的两两相关关系，为了叙述的方便，先对如下的假设条件进行讨论。

设随机向量 $\boldsymbol{x}=(x_1,\boldsymbol{y}')'$，其中 $\boldsymbol{y}=(x_2,\cdots,x_p)'$ 是 $\boldsymbol{x}$ 的后 $p-1$ 维子向量，$E\boldsymbol{x}=0$ 以及

$$\mathrm{cov}(\boldsymbol{x})=\Sigma=\begin{pmatrix}\sigma_{11} & \boldsymbol{\sigma}_{21}' \\ \boldsymbol{\sigma}_{21} & \Sigma_{22}\end{pmatrix}>0,\ \text{且}\ \Sigma_{22}>0, \tag{3.1}$$

其中 $\boldsymbol{\sigma}_{21}$ 是 $p-1$ 维向量。

首先我们希望找到 $\boldsymbol{y}$ 的一个线性组合 $\boldsymbol{\alpha}'\boldsymbol{y}$ 来估计 x_1，其中 $\boldsymbol{\alpha}$ 是 $p-1$ 维常数向量，使得

$$E(x_1-\boldsymbol{\alpha}'\boldsymbol{y})^2 \tag{3.2}$$

达到最小，即 $\boldsymbol{\alpha}'\boldsymbol{y}$ 与 x_1 的差的方差达到最小。
由于

$$\begin{aligned}E(x_1-\boldsymbol{\alpha}'\boldsymbol{y})^2&=\sigma_{11}-2\boldsymbol{\alpha}'\boldsymbol{\sigma}_{21}+\boldsymbol{\alpha}'\Sigma_{22}\boldsymbol{\alpha}\\&=(\sigma_{11}-\boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21})+(\boldsymbol{\alpha}-\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21})'\Sigma_{22}(\boldsymbol{\alpha}\\&\quad-\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21})\end{aligned} \tag{3.3}$$

显然，当且仅当

$$\boldsymbol{\alpha}=\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}\triangleq\boldsymbol{\beta} \tag{3.4}$$

时 $E(x_1-\boldsymbol{\beta}'\boldsymbol{y})^2$ 达到最小值 $\sigma_{1\cdot2}=\sigma_{11}-\boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$

因此可以把 $\boldsymbol{\beta}'\boldsymbol{y}=\boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\boldsymbol{y}$ 作为 x_1 的最佳线性估计，它使得(3.2)式达到最小。

定理 3.1 在上述假设下，x_1 与 $\boldsymbol{\alpha}'\boldsymbol{y}$ 的最大相关在 $\boldsymbol{\beta}'\boldsymbol{y}=\boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\boldsymbol{y}$ 处达到。

证明　由上面所述，对于任意常数 c 及任意的 $p-1$ 维向量 $\boldsymbol{\alpha}$ 总有

$$E(x_1-c\boldsymbol{\alpha}'\boldsymbol{y})^2\geqslant E(x_1-\boldsymbol{\beta}'\boldsymbol{y})^2 \tag{3.5}$$

展开(3.5)式后得到

$$-2c\boldsymbol{\alpha}'\boldsymbol{\sigma}_{21}+c^2\boldsymbol{\alpha}'\Sigma_{22}\boldsymbol{\alpha}\geqslant-2\boldsymbol{\beta}'\boldsymbol{\sigma}_{21}+\boldsymbol{\beta}'\Sigma_{22}\boldsymbol{\beta}$$

选取正数 $c=[\boldsymbol{\beta}'\Sigma_{22}\boldsymbol{\beta}/\boldsymbol{\alpha}'\Sigma_{22}\boldsymbol{\alpha}]^{\frac{1}{2}}$
于是

$$\boldsymbol{\beta}'\boldsymbol{\sigma}_{21}\geqslant\frac{\boldsymbol{\alpha}'\boldsymbol{\sigma}_{21}}{(\boldsymbol{\alpha}'\Sigma_{22}\boldsymbol{\alpha})^{1/2}}\cdot(\boldsymbol{\beta}'\Sigma_{22}\boldsymbol{\beta})^{\frac{1}{2}} \tag{3.6}$$

按相关系数定义

$$\rho(x_1, \alpha' y) = \frac{\operatorname{cov}(x_1, \alpha' y)}{[\operatorname{var} x_1 \cdot \operatorname{var}(\alpha' y)]^{\frac{1}{2}}} = \frac{\alpha' \sigma_{21}}{\sigma_{11}^{\frac{1}{2}}(\alpha' \Sigma_{22} \alpha)^{\frac{1}{2}}}$$

$$\leqslant \frac{\beta' \sigma_{21}}{\sigma_{11}^{\frac{1}{2}}(\beta' \Sigma_{22} \beta)^{\frac{1}{2}}} = \rho(x_1, \beta' y) \tag{3.7}$$

此式左端是 x_1 与任一线性组合 $\alpha' y$ 的相关系数，右端是 x_1 与 $\beta' y$ 的相关系数，从而当 $\alpha = \beta$ 时达到最大相关。 ▌

我们称 x_1 与一切线性组合 $\alpha' y$ 的最大相关系数 $\rho(x_1, \beta' y)$ 为 x_1 与 y 的复相关系数（多重相关系数或全相关系数）记为 $\rho_{x_1 \cdot y}$ 或者 $\rho_{1 \cdot 2, \cdots, p}$。

从(3.7)式中知道，把 $\beta = \Sigma_{22}^{-1} \sigma_{21}$ 代入后有

$$\rho_{x_1 \cdot y} = \left[\frac{\sigma_{21}' \Sigma_{22}^{-1} \sigma_{21}}{\sigma_{11}}\right]^{\frac{1}{2}} \tag{3.8}$$

它有关系式

$$\rho_{x_1 \cdot y}^2 = 1 - \frac{\sigma_{1 \cdot 2}}{\sigma_{11}} \text{ 或者 } \sigma_{1 \cdot 2} = \sigma_{11}(1 - \rho_{x_1 \cdot y}^2) \tag{3.9}$$

其中 $\sigma_{1 \cdot 2}$ 可看成 x_1 在消去 y 的最大线性效应后的方差。

设 $x = (x_1, x_2, z')'$ 是一 p 维随机向量，z 是 $p-2$ 维子向量，$Ex = 0$ 以及

$$\operatorname{cov}(x) = \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{31}' \\ \sigma_{21} & \sigma_{22} & \sigma_{32}' \\ \sigma_{31} & \sigma_{32} & \Sigma_{33} \end{pmatrix} > 0, \quad (\Sigma_{33} > 0)$$

令 $a'z$ 是 x_1 的最佳线性估计（它使得 x_1 与 $a'z$ 之间有最大相关），$b'z$ 是 x_2 的最佳线性估计，这里 a, b 是 $p-2$ 维常数向量。那么如上可求得

$$a = \Sigma_{33}^{-1} \sigma_{31}, \quad b = \Sigma_{33}^{-1} \sigma_{32}$$

从而消去 z 的最大线性效应后得到的两个随机变数为

$$y_1 = x_1 - \sigma_{31}' \Sigma_{33}^{-1} z, \quad y_2 = x_2 - \sigma_{32}' \Sigma_{33}^{-1} z.$$

定义在 z 给定条件下 x_1 和 x_2 的偏相关系数为上述 y_1 和

y_2 的相关系数，记为 $\rho_{12\cdot z}$ 或者 $\rho_{12\cdot 3,\cdots,p}$。从而

$$\rho_{12\cdot z}=\frac{\operatorname{cov}(x_1-\boldsymbol{\sigma}'_{31}\Sigma_{33}^{-1}\boldsymbol{z},x_2-\boldsymbol{\sigma}'_{32}\Sigma_{33}^{-1}\boldsymbol{z})}{[\operatorname{var}(x_1-\boldsymbol{\sigma}'_{31}\Sigma_{33}^{-1}\boldsymbol{z})\cdot\operatorname{var}(x_2-\boldsymbol{\sigma}'_{32}\Sigma_{33}^{-1}\boldsymbol{z})]^{\frac{1}{2}}}$$

$$=\frac{(\sigma_{12}-\boldsymbol{\sigma}'_{31}\Sigma_{33}^{-1}\boldsymbol{\sigma}_{32})}{(\sigma_{11}-\boldsymbol{\sigma}'_{31}\Sigma_{33}^{-1}\boldsymbol{\sigma}_{31})^{\frac{1}{2}}(\sigma_{22}-\boldsymbol{\sigma}'_{32}\Sigma_{33}^{-1}\boldsymbol{\sigma}_{32})^{\frac{1}{2}}}\tag{3.10}$$

上面为了使读者能深刻理解各种相关关系，均以最简单的形式给出有关表达式，而在实际运用中只要略作变化即可得到一些形式上复杂的表达式。例如若 $\boldsymbol{x}=(x_1,\cdots,x_{10})'$ 那么 $\rho_{2\cdot 13,\cdots,10}$ 表示 x_2 与其余分量间的复相关系数。又如 $\rho_{25\cdot 6,\cdots,10}$ 表示在 $\boldsymbol{z}=(x_6,\cdots,x_{10})'$ 给定下 x_2 与 x_5 的偏相关系数，为此只须取出 Σ 中与 $x_2,x_5,\boldsymbol{z}$ 有关的诸行，诸列得到新的协方差矩阵，然后按(3.10)式计算得到。

（二） 正态向量的条件分布

本段将着重讨论正态向量的条件分布及有关概念。

定理 3.2 设 p 维随机向量 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\Sigma)$，其中

$$\boldsymbol{x}=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{pmatrix},\ \boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix},\ \Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\geqslant 0,\ \ \Sigma_{22}>0,$$

$\boldsymbol{x}_1$ 是 r 维，$\boldsymbol{x}_2$ 是 s 维 $(s+r=p)$，$\boldsymbol{\mu}$ 及 Σ 相应分块。那么在给定 $\boldsymbol{x}_2$ 后 $\boldsymbol{x}_1$ 的条件分布为

$$\boldsymbol{x}_1|\boldsymbol{x}_2\sim N_r(\boldsymbol{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2),\Sigma_{1\cdot 2})$$

即给定 $\boldsymbol{x}_2$ 后 $\boldsymbol{x}_1$ 的条件分布为 r 维正态分布且

$$E(\boldsymbol{x}_1|\boldsymbol{x}_2)=\boldsymbol{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)\tag{3.11}$$

$$\operatorname{cov}(\boldsymbol{x}_1|\boldsymbol{x}_2)=\Sigma_{1\cdot 2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\tag{3.12}$$

证明 作非奇异线性变换

$$\boldsymbol{y}=\begin{pmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{pmatrix}=\begin{pmatrix}I & -\Sigma_{12}\Sigma_{22}^{-1}\\ 0 & I\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{x}_1-\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{x}_2\\ \boldsymbol{x}_2\end{pmatrix}$$

由定理 2.2 知

$$\boldsymbol{y}=\begin{pmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{pmatrix}\sim N_p\left(\begin{pmatrix}\boldsymbol{\mu}_1-\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2\\ \boldsymbol{\mu}_2\end{pmatrix},\ \begin{pmatrix}\Sigma_{1\cdot 2} & 0\\ 0 & \Sigma_{22}\end{pmatrix}\right)$$

由定理 2.5 知 $\boldsymbol{y}_1$ 与 $\boldsymbol{y}_2$ 相互独立且

$$\boldsymbol{y}_1=\boldsymbol{x}_1-\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{x}_2\sim N_r(\boldsymbol{\mu}_1-\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2,\ \Sigma_{1\cdot2})$$

从而当 $\boldsymbol{x}_2$ 给定后

$$\boldsymbol{x}_1\sim N_r(\boldsymbol{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2),\Sigma_{1\cdot2}).\qquad\blacksquare$$

此时我们称方程

$$\boldsymbol{x}_1=E(\boldsymbol{x}_1|\boldsymbol{x}_2)=\boldsymbol{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)$$

为 $\boldsymbol{x}_1$ 关于 $\boldsymbol{x}_2$ 的回归曲面(或回归方程),而系数 $\Sigma_{12}\Sigma_{22}^{-1}$ 称为 $\boldsymbol{x}_1$ 关于 $\boldsymbol{x}_2$ 的回归系数矩阵, $\Sigma_{1\cdot2}$ 称为 $\boldsymbol{x}_1$ 在 $\boldsymbol{x}_2$ 给定下的条件协方差矩阵。

若 $r=1$, 此时 x_1 的最佳线性估计正是 x_1 关于 $\boldsymbol{x}_2$ 的回归平面。

若记 $\Sigma_{1\cdot2}=(\sigma_{ij\cdot r+1,\cdots,p})$, 它是 $r\times r$ 阶矩阵,则称

$$\rho_{ij\cdot r+1,\cdots,p}=\frac{\sigma_{ij\cdot r+1,\cdots,p}}{\sqrt{\sigma_{ii\cdot r+1,\cdots,p}\cdot\sigma_{jj\cdot r+1,\cdots,p}}}$$

为给定 $\boldsymbol{x}_2$ 后 $\boldsymbol{x}_1$ 的分量 x_i 和 x_j 的条件相关系数 (i, $j=1,2,\cdots,r$)。与(3.10)式比较,可以知道,当 $\boldsymbol{x}$ 服从正态分布时,在 $\boldsymbol{x}_2$ 给定条件下 $\boldsymbol{x}_1$ 诸分量的偏相关系数正是它的条件相关系数,为此只须记

$$\Sigma=\begin{pmatrix}\sigma_{11}&\sigma_{12}&\boldsymbol{\sigma}_{31}'\\\sigma_{21}&\sigma_{22}&\boldsymbol{\sigma}_{32}'\\\boldsymbol{\sigma}_{31}&\boldsymbol{\sigma}_{32}&\Sigma_{33}\end{pmatrix}\quad\Sigma_{11}=\begin{pmatrix}\sigma_{11}&\sigma_{12}\\\sigma_{21}&\sigma_{22}\end{pmatrix}$$

于是

$$\begin{aligned}\Sigma_{1\cdot3}&=\begin{pmatrix}\sigma_{11}&\sigma_{12}\\\sigma_{21}&\sigma_{22}\end{pmatrix}-\begin{pmatrix}\boldsymbol{\sigma}_{31}'\\\boldsymbol{\sigma}_{32}'\end{pmatrix}\Sigma_{33}^{-1}(\boldsymbol{\sigma}_{31},\boldsymbol{\sigma}_{32})\\&=\begin{pmatrix}\sigma_{11}-\boldsymbol{\sigma}_{31}'\Sigma_{33}^{-1}\boldsymbol{\sigma}_{31}&\sigma_{12}-\boldsymbol{\sigma}_{31}'\Sigma_{33}^{-1}\boldsymbol{\sigma}_{32}\\\sigma_{21}-\boldsymbol{\sigma}_{32}'\Sigma_{33}^{-1}\boldsymbol{\sigma}_{31}&\sigma_{22}-\boldsymbol{\sigma}_{32}'\Sigma_{33}^{-1}\boldsymbol{\sigma}_{32}\end{pmatrix}\\&=\mathrm{cov}\left(\begin{pmatrix}x_1\\x_2\end{pmatrix}\middle|\boldsymbol{z}\right)\end{aligned}$$

即可看出。

注意上述结论"偏相关系数等于条件相关系数"仅对正态分布成立，其他分布时未必成立。

另外若 $r=1$，由定理 3.2 知道在 $\boldsymbol{x}_2$ 给定下 x_1 的条件方差为

$$\sigma_{1\cdot 2}=\sigma_{11}-\boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$$

与(3.9)式 $\sigma_{1\cdot 2}=\sigma_{11}(1-\rho_{1\cdot 2,\cdots,p}^2)$

相对照，说明 x_1 在 $\boldsymbol{x}_2$ 给定后的条件方差小于它自身的原来方差 σ_{11}，且若 x_1 与 $\boldsymbol{x}_2$ 的复相关系数绝对值越接近于 1 即 x_1 与 $\boldsymbol{x}_2$ 的线性相关程度越大，则条件方差越小。

例 3.1　在例 2.1 中

$$\begin{pmatrix}x_1\\x_2\end{pmatrix}\sim N_2\left(\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix},\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$$

那么在 x_2 给定条件下 x_1 的条件分布为

$$x_1|x_2\sim N\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2),\sigma_1^2(1-\rho^2)\right)$$

$\sigma_{1\cdot 2}=\sigma_1^2(1-\rho^2)$ 正是条件方差，x_1 关于 x_2 的回归直线为

$$x_1=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2).$$

（三）　偏相关系数的递推公式

根据(3.10)式我们可以求出偏相关系数，然而十分不便，我们希望找到一个递推公式，能够以简单的相关系数 ρ_{ij} 出发求出所需要的偏相关系数。为此我们介绍下面定理，其证明将在本章的附录中给出。

定理 3.3　偏相关系数存在下列递推公式

$$\rho_{ij\cdot q,\cdots,p}=\frac{\rho_{ij\cdot q+1,\cdots,p}-\rho_{iq\cdot q+1,\cdots,p}\cdot\rho_{jq\cdot q+1,\cdots,p}}{\sqrt{1-\rho_{iq\cdot q+1,\cdots,p}^2}\cdot\sqrt{1-\rho_{jq\cdot q+1,\cdots,p}^2}}\qquad(3.13)$$

由这个公式可以根据 $\{\rho_{ij}\}$ 计算出 $\{\rho_{ij\cdot p}\}$ 再计算出 $\{\rho_{ij\cdot p-1,p}\},\cdots,\{\rho_{12\cdot 3,\cdots,p}\}$ 等等。常用的有

$$\rho_{12\cdot3}=\frac{\rho_{12}-\rho_{13}\rho_{23}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}} \tag{3.14}$$

因此只要知道相关矩阵，那么各种偏相关系数也就完全确定了。

例 3.2 在制订服装标准时，需抽样进行人体测量，今从女子身体测量中取出部分结果如下：x_1：身高 x_2：胸围，x_3：腰围，x_4：上体长，x_5：臀围。经计算知道这五个变量的相关矩阵为

$$R=\begin{pmatrix}1.000 & & & & \\ 0.216 & 1.000 & & & \\ 0.054 & 0.732 & 1.000 & & \\ 0.648 & 0.242 & 0.133 & 1.000 & \\ 0.368 & 0.676 & 0.627 & 0.376 & 1.000\end{pmatrix}$$

首先可以看到 $\rho_{14}=0.648$，$\rho_{23}=0.732$，这说明上体长与身高，胸围与腰围之间存在较大的正相关，而 $\rho_{13}=0.054$ 说明身高与腰围仅存在极微弱的正相关，几乎是互不相关的，下面我们计算偏相关系数 $\rho_{34\cdot5}$ 和 $\rho_{23\cdot1}$。

$$\rho_{34\cdot5}=[0.133-0.627\times0.376]/\sqrt{1-(0.627)^2}\cdot\sqrt{1-(0.376)^2}=-0.1423,$$

$$\rho_{23\cdot1}=[0.732-0.216\times0.054]/\sqrt{1-(0.216)^2}\cdot\sqrt{1-(0.054)^2}=0.7388.$$

比较 ρ_{23} 和 $\rho_{23\cdot1}$ 知道腰围和胸围即使排除了身高的影响后仍高度正相关。然而从 ρ_{34} 和 $\rho_{34\cdot5}$ 的比较发现，把臀围的影响去除后上体长与腰围变成了负相关。这正说明了女性主要特征臀围造成的变化，因此我们从变量的相关系数说明变量的相互关系时要十分慎重。

§2.4 正态随机矩阵的若干性质

本节着重讨论正态随机矩阵的一些性质。

设 X 是 $p\times n$ 阶随机矩阵，M 是 $p\times n$ 阶常数矩阵，Σ 是 $p\times$

p 阶非负定矩阵，V 是 $n\times n$ 阶非负定矩阵，且记

$$X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)=(\boldsymbol{x}_{(1)},\cdots,\boldsymbol{x}_{(p)})'$$

$$M=(\boldsymbol{\mu}_1,\cdots,\boldsymbol{\mu}_n)=(\boldsymbol{\mu}_{(1)},\cdots,\boldsymbol{\mu}_{(p)})'$$

定义 4.1 如果 $\mathrm{vec}X\sim N_{pn}(\mathrm{vec}M,\ \Sigma\otimes V)$，则称随机矩阵 X 服从正态分布 $N_{p,n}(M,\Sigma,V)$

若 $M=0$，$\Sigma=I_p$，$V=I_n$，则称 X 服从标准正态分布 $N_{p,n}(0,I_p,I_n)$。此时 X 的各元素都相互独立且 $\sim N(0,1)$。

如果 $\mathrm{rk}(\Sigma)=r$，$\mathrm{rk}(V)=s$ $(0<r\leqslant p, 0<s\leqslant n)$ 那么存在秩为 r 的 $p\times r$ 阶矩阵 A 和秩为 s 的 $n\times s$ 阶矩阵 B，使得 $\Sigma=AA'$ 和 $V=BB'$ 类似于定义 2.4，我们可以给出另一等价定义：

定义 4.2 若 $p\times n$ 阶随机矩阵 X 可以表示成

$$X=M+AYB'$$

其中 Y 是 $r\times s$ 阶随机矩阵 $\sim N_{r,s}(0,I_r,I_s)$，A 是秩为 r 的 $p\times r$ 阶矩阵，B 是秩为 s 的 $n\times s$ 阶矩阵，M 是 $p\times n$ 阶常数矩阵，令 $\Sigma=AA'$，$V=BB'$，那么称 $X\sim N_{p,n}(M,\Sigma,V)$。

从定义中我们可以看出下列几个性质：

1）列向量 $\boldsymbol{x}_j\sim N_p(\boldsymbol{\mu}_j, v_{jj}\Sigma)$

2）行向量 $\boldsymbol{x}_{(i)}\sim N_n(\boldsymbol{\mu}_{(i)},\sigma_{ii}V)$

3）$\mathrm{cov}(\boldsymbol{x}_i,\boldsymbol{x}_j)=v_{ij}\Sigma$

4）$\mathrm{cov}(\boldsymbol{x}_{(i)},\boldsymbol{x}_{(j)})=\sigma_{ij}V$

5）$\mathrm{cov}(x_{ij},x_{kl})=\sigma_{ik}v_{jl}$

6）$X'\sim N_{n,p}(M',V,\Sigma)$

其中 $\Sigma=(\sigma_{ij})_{p\times p}$，$V=(v_{ij})_{n\times n}$。

由此可看出 X 的诸行(列)都是正态分布着的，但不一定是相互独立。

当 Σ 和 V 都是正定矩阵时，$\mathrm{vec}X$ 的密度可写成

$$f(\mathrm{vec}X)=(2\pi)^{-\frac{pn}{2}}|\Sigma\otimes V|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\mathrm{vec}X\right.$$

$$-\text{vec}\mu)'(\Sigma\otimes V)^{-1}(\text{vec}X-\text{vec}\mu)\Big\}\quad(4.1)$$

利用 $|\Sigma\otimes V|=|\Sigma|^n|V|^p$ 以及 $\text{vec}(AXB)=(A\otimes B')\text{vec}X$

$$\text{vec}X'\text{vec}Y=\text{tr}X'Y,\ (\Sigma\otimes V)^{-1}=\Sigma^{-1}\otimes V^{-1}.$$

那么我们可把(4.1)改写成

$$f(X)=(2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}|V|^{-\frac{p}{2}}\text{etr}\left\{-\frac{1}{2}\Sigma^{-1}(X-M)V^{-1}\cdot(X-M)'\right\}\quad(4.2)$$

其中 $\text{etr}(\cdot)$ 表示 $\exp\{\text{tr}(\cdot)\}$.

在实际问题中经常假定被研究的对象服从多元正态分布，但参数 $\boldsymbol{\mu}$ 及 Σ 是未知的，这时我们必须从总体 $\sim N_p(\boldsymbol{\mu},\Sigma)$ 中抽取 n 个独立样本 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 来估计 $\boldsymbol{\mu}$ 及 Σ. 这时 $X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)$ 被称为观察值矩阵或样本资料矩阵. 利用上面记号，若再记

$M=(\boldsymbol{\mu},\boldsymbol{\mu},\cdots,\boldsymbol{\mu})=\boldsymbol{\mu}\cdot\mathbf{1}_n'$

($\mathbf{1}_n$ 表示元素全为 1 的 n 维列向量)，那么可简记为

$$X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)\sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}_n',\Sigma,I_n)$$

而 X 的密度可写成

$$f(X)=(2\pi)^{-\frac{pn}{2}}|\Sigma|^{-\frac{n}{2}}\text{etr}\left\{-\frac{1}{2}\Sigma^{-1}(X-\boldsymbol{\mu}\mathbf{1}_n')(X-\boldsymbol{\mu}\mathbf{1}_n')'\right\}\quad(4.3)$$

以后在讨论参数估计和假设检验时，总用上述记号表示样本随机矩阵.

下面我们讨论随机矩阵 $X\sim N_{p,n}(M,\Sigma,V)$ 的若干性质.

定理 4.1 如果 $X\sim N_{p,n}(M,\Sigma,V)$，$A$ 是 $m\times p$ 阶矩阵，B 是 $n\times s$ 阶矩阵，则

$$AXB\sim N_{m,s}(AMB,A\Sigma A',B'VB)$$

证明 此时 $\text{vec}X\sim N_{p,n}(\text{vec}M,\Sigma\otimes V)$

所以

$$\mathrm{vec}(AXB)=(A\otimes B')\mathrm{vec}X\sim N_{m,s}((A\otimes B')\mathrm{vec}M,\ (A\otimes B')(\Sigma\otimes V)(A'\otimes B))$$

即

$$\mathrm{vec}(AXB)\sim N_{m,s}(\mathrm{vec}(AMB),A\Sigma A'\otimes B'VB)$$

按定义得

$$AXB\sim N_{m,s}(AMB,A\Sigma A',B'VB).\ \blacksquare$$

对于X的子矩阵的分布，只须在定理 4.1 中取 A,B 为特殊矩阵即可得到。

例如，若记 $X\sim N_{p,n}(M,\Sigma,V)$，其中

$$X=\begin{pmatrix}X_{11} & X_{12}\\ X_{21} & X_{22}\end{pmatrix},\quad M=\begin{pmatrix}M_{11} & M_{12}\\ M_{21} & M_{22}\end{pmatrix},\quad \Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix},$$

$$V=\begin{pmatrix}V_{11} & V_{12}\\ V_{21} & V_{22}\end{pmatrix},$$

其中 X_{11},M_{11} 是 $r\times s$ 阶矩阵，Σ_{11} 是 $r\times r$ 阶非负定矩阵，V_{11} 是 $s\times s$ 阶非负定矩阵，那么 $X_{11}\sim N_{r,s}(M_{11},\Sigma_{11},V_{11})$。此时相当于取$A$为 $r\times p$ 阶矩阵 $(I_r,0)$，B 取 $n\times s$ 阶矩阵 $\begin{pmatrix}I_s\\ 0\end{pmatrix}$ 即可。

定理 4.2 设 $X\sim N_{p,n}(M,\Sigma,V)$，$A$ 是 $n\times r$ 阶矩阵，B 是 $n\times s$ 阶矩阵。则 $Y=XA$ 和 $Z=XB$ 相互独立的充分必要条件是 $A'VB=0$。

证明　只须证明 $\mathrm{vec}X$ 和 $\mathrm{vec}Y$ 相互独立的充分必要条件是 $A'VB=0$。由定理条件知道

$$\mathrm{vec}Y=(I\otimes A')\mathrm{vec}X\sim N_{p,r}(\mathrm{vec}MA,\Sigma\otimes A'VA)$$

$$\mathrm{vec}Z=(I\otimes B')\mathrm{vec}X\sim N_{p,s}(\mathrm{vec}MB,\Sigma\otimes B'VB).$$

$\mathrm{vec}Y$ 和 $\mathrm{vec}Z$ 相互独立的充分必要条件是

$$\begin{aligned}\mathrm{cov}(\mathrm{vec}Y,\mathrm{vec}Z)&=\mathrm{cov}((I\otimes A')\mathrm{vec}X,(I\otimes B')\mathrm{vec}X)\\&=(I\otimes A')(\Sigma\otimes V)(I\otimes B')'=\Sigma\otimes A'VB\\&=0.\end{aligned}$$

由于Σ不恒为零，即为 $A'VB=0$。$\blacksquare$

类似地可以证明：

定理 4.3 设 $X \sim N_{p,n}(M,\Sigma,V)$ A 是 $m \times p$ 阶矩阵，B 是 $s \times p$ 阶矩阵，则 $Y = AX$ 和 $Z = BX$ 相互独立的充分必要条件是 $A\Sigma B' = 0$。

下面我们求 $X \sim N_{p,n}(M,\Sigma,V)$ 相应的特征函数。

设 $T = (t_{ij})$ 是任意的 $p \times n$ 阶实数矩阵，则 $\mathrm{vec}X$ 的特征函数为

$$\exp\left\{i\mathrm{vec}T'\mathrm{vec}M - \frac{1}{2}\mathrm{vec}T'(\Sigma \otimes V)\mathrm{vec}T\right\}$$

由于

$$\mathrm{vec}T'\mathrm{vec}M = \mathrm{tr}T'M,$$

$$\mathrm{vec}T'(\Sigma \otimes V)\mathrm{vec}T = \mathrm{vec}T'\mathrm{vec}(\Sigma TV) = \mathrm{tr}T'\Sigma TV,$$

所以 X 的特征函数为

$$\varphi_X(T) = E[\exp(i\mathrm{tr}T'X)] = \exp\left\{i\mathrm{tr}T'M - \frac{1}{2}\mathrm{tr}T'\Sigma TV\right\} \tag{4.4}$$

与随机向量一样，正态随机矩阵也可用它的特征函数来等价地定义。读者可自行推出正态随机矩阵的各种性质。

习 题 二

2.1 已知二维随机向量 $(x,y)'$ 有密度

$$f(x,y) = \begin{cases} 1, & 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1 \\ 0, & \text{其他} \end{cases}$$

试求 1) 联合分布函数 $F(x,y)$，

2) 边缘分布函数 $F_i(x)$ 和边缘密度 $f_i(x)$，$i = 1,2$。

3) 条件分布密度 $f(x|y)$（注意若 $f(x_0,y_0)=0$，则定义 $f(x_0|y_0) = 0$）。

4) 证明 x 和 y 独立。

2.2 设 x 的分布函数 $F_1(x)$ 在 x_0 处连续，y 的分布函数 $F_2(y)$ 在 y_0 连续，试证明 (x,y) 的联合分布函数 $F(x,y)$ 在 (x_0,y_0) 处连续。（这

里假定 $F(x,y)$ 存在)

2.3 设 $x=(x_1,x_2)'$ 有特征函数

$$\varphi(t_1,t_2)=a[e^{i(t_1+t_2)}+1]+be^{it_1}+be^{it_2},$$

其中 a,b 是满足 $a+b=\frac{1}{2}$ 的正常数，求 x 的协方差矩阵.

提示：利用特征函数性质 4).

2.4 利用特征函数证明 当 $x\sim N_p(0,\Sigma)$(其中 $x=(x_1,\cdots,x_p)'$，$\Sigma=(\sigma_{ij})_{p\times p}$) 时，有

1) $Ex_ix_jx_k=0$

2) $Ex_ix_jx_kx_l=\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}$

2.5 设 x_1,x_2,x_3 分别是三个连续型随机变量，x_1 和 x_2 有二阶矩存在，证明

1) $Ex_1=E[E(x_1|x_2)]$

2) $\operatorname{var}x_1=E[\operatorname{var}(x_1|x_2)]+\operatorname{var}[E(x_1|x_2)]$

3) $\operatorname{cov}(x_1,x_2)=E[\operatorname{cov}(x_1,x_2|x_3)+\operatorname{cov}[E(x_1|x_3),E(x_2|x_3)]$.

提示：利用方差和协方差的定义.

2.6 设 $x=(x_1,x_2)'\sim N_2(0,\Sigma)$，$\Sigma=\begin{pmatrix}4 & 2\\ 2 & 1\end{pmatrix}$，试求出一个矩阵 A，使得 $x=Ay$，且 y 是非奇异正态分布，并给出 y 的分布密度.

2.7 设 $x=(x_1,x_2)'$ 有密度

$$f(x_1,x_2)=\frac{1}{2\pi}\exp\left\{-\frac{1}{2}(2x_1^2+x_2^2+2x_1x_2-22x_1-14x_2+65)\right\}$$

求：$\mu=Ex$ 和 $\Sigma=\operatorname{cov}(x)$.

提示：将 $f(x_1,x_2)$ 写成正态分布密度的矩阵形式(2.3).

2.8 设 $x=(x_1,x_2)'$ 有密度

$$f(x_1,x_2)=\begin{cases}\dfrac{1}{\pi}\exp\left\{-\dfrac{x_1^2+x_2^2}{2}\right\}, & x_1x_2>0,\\ 0, & \text{其他},\end{cases}$$

证明 x_1 和 x_2 的边缘分布是正态分布，但 x 不是二元正态分布.

2.9 已知 $x=(x_1,x_2,x_3)'$ 的协方差矩阵为

$$\operatorname{cov}(x)=\Sigma=\begin{pmatrix}5 & 2 & 3\\ 2 & 3 & 0\\ 3 & 0 & 2\end{pmatrix},$$

1) 求 $\mathrm{var}(x_1-2x_2+x_3)$。

2) 设 $y_1=x_1+x_2$, $y_2=x_1+x_2+x_3$, 求 $\mathrm{cov}(y_1,y_2)$。

2.10 已知 x_1,x_2 独立且服从相同分布 $N(\mu,\sigma^2)$。求 $y_1=\alpha x_1+\beta x_2$ 和 $y_2=\alpha x_1-\beta x_2$ 的相关系数。

2.11 设 x 与诸 $y_i(i=1,\cdots,n)$ 都不相关，证明对于任意 $\alpha_i(i=1,\cdots,n)$，x 与 $\sum_{i=1}^{n}\alpha_i y_i$ 不相关。

2.12 设 $\boldsymbol{x}=(x_1,x_2,x_3)'\sim N_3(\boldsymbol{0},I_3)$，求

$$y=\frac{x_1+x_2x_3}{\sqrt{1+x_3^2}}$$

的分布密度。

提示：考虑在给定 x_3 条件下 y 的分布，它是正态变量的线性组合，求出 $f(x_3,y)$ 后再求 y 的分布密度。

2.13 证明特征函数的性质 1，2，5。

2.14 设 $\boldsymbol{x}=(x_1,x_2)'\sim N_2(0,I_2)$，求在 x_1+x_2 给定下 x_1 的条件分布。

提示：先求 x_1,x_1+x_2 的联合分布，再求条件分布。

2.15 设 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\Sigma)$，其中 $\boldsymbol{\mu}=\mu_0\cdot\mathbf{1}_p$，$\Sigma=\sigma^2\cdot I_p$，$A=I_p-\frac{1}{p}\mathbf{1}_p\mathbf{1}_p'$，求证

$$E(\boldsymbol{x}'A\boldsymbol{x})=(p-1)\sigma^2.$$

提示：利用定理 1.5。

*)2.16 设 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\Sigma)$，$\Sigma>0$，$\boldsymbol{x}=(x_1,\cdots,x_p)'$，定义 $\log\boldsymbol{y}=(\log y_1,\cdots,\log y_p)'$ 若 $x_i=\log y_i$ 有上述正态分布，则称 $\boldsymbol{y}$ 服从 p 维对数正态分布，证明它的密度为

$$f(\boldsymbol{y})=\begin{cases}(2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}\left(\prod_{i=1}^{p}y_i\right)^{-1}\exp\left\{-\frac{1}{2}(\log\boldsymbol{y}-\boldsymbol{\mu})'\Sigma^{-1}(\log\boldsymbol{y}-\boldsymbol{\mu})\right\}, & (y_i>0,\ i=1,\cdots,p),\\ 0, & \text{其他},\end{cases}$$

并证明

1) $Ey_i^r=\exp\left\{r\mu_i+\frac{1}{2}r^2\sigma_{ii}\right\}$

2) $\operatorname{var}(y_i)=\exp\{2\mu_i+2\sigma_{ii}\}-\exp\{2\mu_i+\sigma_{ii}\}$。

附　　录

在这个附录里我们将证明定理 3.3“偏相关系数的递推公式”。 由于正态分布条件相关系数等于偏相关系数，我们只须推导正态分布的条件相关系数的递推公式。为此我们先证明一个引理。

引理 3.1　设 p 维随机向量 $\boldsymbol{x}\sim N_p(\mathbf{0},\Sigma)$，其中

$$\boldsymbol{x}=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\\ \boldsymbol{x}_3\end{pmatrix}\begin{matrix}p_1\\ p_2\\ p_3\end{matrix}\qquad \Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12} & \Sigma_{13}\\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23}\\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33}\end{pmatrix}\qquad \sum_{i=1}^{3}p_i=p$$

$\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3$ 分别为 p_1,p_2,p_3 维子向量。那么有

$$\Sigma_{11}-(\Sigma_{12},\Sigma_{13})\begin{pmatrix}\Sigma_{22} & \Sigma_{23}\\ \Sigma_{32} & \Sigma_{33}\end{pmatrix}^{-1}\begin{pmatrix}\Sigma_{12}\\ \Sigma_{13}\end{pmatrix}=\Sigma_{11\cdot3}-\Sigma_{12\cdot3}\Sigma_{22\cdot3}^{-1}\Sigma_{21\cdot3}\qquad (1)$$

这里

$$\Sigma_{ij\cdot3}=\Sigma_{ij}-\Sigma_{i3}\Sigma_{33}^{-1}\Sigma_{3j}\quad (i,j=1,2)$$

证明　记 $\boldsymbol{x}$ 的分布密度为 $f(\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \boldsymbol{x}_3)$，相应于 $\boldsymbol{x}_i$ 的边缘密度为 $f(\boldsymbol{x}_i)$。在 $\boldsymbol{x}_i$ 给定下 $\boldsymbol{x}_j$ 的条件密度为 $f(\boldsymbol{x}_j|\boldsymbol{x}_i)$。于是

$$f(\boldsymbol{x}_1|\boldsymbol{x}_2,\boldsymbol{x}_3)=\frac{f(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3)}{f(\boldsymbol{x}_2,\boldsymbol{x}_3)}=\frac{f(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3)/f(\boldsymbol{x}_3)}{f(\boldsymbol{x}_2,\ \boldsymbol{x}_3)/f(\boldsymbol{x}_3)}$$

$$=\frac{f(\boldsymbol{x}_1,\boldsymbol{x}_2|\boldsymbol{x}_3)}{f(\boldsymbol{x}_2|\boldsymbol{x}_3)}=f(\boldsymbol{x}_1|\boldsymbol{x}_2\ \big|\boldsymbol{x}_3)。$$

因此

$$\operatorname{cov}(\boldsymbol{x}_1|\boldsymbol{x}_2,\boldsymbol{x}_3)=\operatorname{cov}(\boldsymbol{x}_1|\boldsymbol{x}_2\big|\boldsymbol{x}_3)$$

而 $\operatorname{cov}(\boldsymbol{x}_1|\boldsymbol{x}_2,\boldsymbol{x}_3)$ 正是在 $(\boldsymbol{x}_2,\ \boldsymbol{x}_3)$ 给定条件下 $\boldsymbol{x}_1$ 的条件协方差矩阵，即(1)式的左端表达式。

另外在 $\boldsymbol{x}_3$ 给定后 $(\boldsymbol{x}_1,\boldsymbol{x}_2)$ 的条件协方差矩阵为

$$\operatorname{cov}\left(\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{pmatrix}\Big|\ \boldsymbol{x}_3\right)=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}-\begin{pmatrix}\Sigma_{13}\\ \Sigma_{23}\end{pmatrix}\Sigma_{33}^{-1}(\Sigma_{31},\ \Sigma_{32})$$

$$=\begin{pmatrix}\Sigma_{11\cdot3} & \Sigma_{12\cdot3}\\ \Sigma_{21\cdot3} & \Sigma_{22\cdot3}\end{pmatrix}$$

此时 $\boldsymbol{x}_1$ 关于 $\boldsymbol{x}_2$ 给定下的条件协方差矩阵为

$$\mathrm{cov}(x_1|x_2\Big|x_3) = \Sigma_{11\cdot3} - \Sigma_{12\cdot3}\Sigma_{22\cdot3}^{-1}\Sigma_{21\cdot3}$$

与(1)式右端相同，引理得证。 ▌

引理 3.1 告诉我们，x_1 在 (x_2, x_3) 给定下的条件协方差矩阵与先让 x_3 给定再求 x_2 给定下 x_1 的条件协方差矩阵相同，即 x_1 的条件协方差矩阵与 x_2, x_3 的给定次序无关。

定理 3.3 的证明：

在引理 3.1 中令 $p_1 = q-1$，$p_2 = 1$，$p_3 = p-q$。于是(1)式中左端和右端各矩阵的元素分别为

$$\mathrm{cov}(x_1|x_q, x_3) = (\sigma_{ij\cdot q,\dots,p}) \qquad \Sigma_{11\cdot3} = (\sigma_{ij\cdot q+1,\dots,p})$$

$$\Sigma_{12\cdot3} = (\sigma_{iq\cdot q+1,\dots,p}) \qquad \Sigma_{22\cdot3} = \sigma_{qq\cdot q+1,\dots,p}$$

从而(1)按元素写出正是

$$\sigma_{ij\cdot q,\dots,p} = \sigma_{ij\cdot q+1,\dots,p} - \frac{\sigma_{iq\cdot q+1,\dots,p}\cdot\sigma_{jq\cdot q+1,\dots,p}}{\sigma_{qq\cdot q+1,\dots,p}} \tag{2}$$

在(2)式中令 $i = j$，

$$\sigma_{ii\cdot q,\dots,p} = \sigma_{ii\cdot q+1,\dots,p}\left(1 - \frac{\sigma^2_{iq\cdot q+1,\dots,p}}{\sigma_{qq\cdot q+1,\dots,p}\cdot\sigma_{ii\cdot q+1,\dots,p}}\right) \tag{3}$$

$$= \sigma_{ii\cdot q+1,\dots,p}(1 - \rho^2_{iq\cdot q+1,\dots,p})$$

而 $$\rho_{ij\cdot q,\dots,p} = \sigma_{ij\cdot q,\dots,p}\Big/\sqrt{\sigma_{ii\cdot q,\dots,p}\cdot\sigma_{jj\cdot q,\dots,p}} \tag{4}$$

将(2)和(3)式代入(4)式之中得到

$$\rho_{ij\cdot q,\dots,p} = \frac{\sigma_{ij\cdot q+1,\dots,p} - \dfrac{\sigma_{iq\cdot q+1,\dots,p}\sigma_{jq\cdot q+1,\dots,p}}{\sigma_{qq\cdot q+1,\dots,p}}}{\sqrt{\sigma_{ii\cdot q+1,\dots,p}(1-\rho^2_{iq\cdot q+1,\dots,p})}\sqrt{\sigma_{jj\cdot q+1,\dots,p}(1-\rho^2_{jq\cdot q+1,\dots,p})}}$$

$$= \frac{\rho_{ij\cdot q+1,\dots,p} - \rho_{iq\cdot q+1,\dots,p}\cdot\rho_{jq\cdot q+1,\dots,p}}{\sqrt{1-\rho^2_{iq\cdot q+1,\dots,p}}\sqrt{1-\rho^2_{jq\cdot q+1,\dots,p}}}.$$ ▌

第三章　Wishart 分布和 T^2- 统计量，Λ- 统计量的分布

在本章中我们将讨论与多元正态分布有着密切关系的若干重要的分布，其中 Wishart 分布是一元统计分析中 χ^2 分布的自然推广．Hotelling T^2 分布是一元 t 分布的自然推广，而 Wilks Λ 统计量的分布与一元 F 分布有着密切的关系．为了与一元统计中诸分布进行比较，我们将在第一节中较详细地探讨关于二次型分布的一些结论。

§ 3.1　关于二次型分布的一些结论

本节首先回顾一下一元统计中常用的 χ^2 分布，t 分布和 F 分布的定义，然后对二次型服从 χ^2 分布及其二次型相互独立的充分必要条件进行讨论。

（一）χ^2 分布，t 分布，F 分布，β 分布

（1）χ^2 分布

设 p 维随机向量 $\boldsymbol{x}=(x_1,\cdots,x_p)'\sim N_p(\boldsymbol{\mu},I_p)$，我们称随机变量 $y=\boldsymbol{x}'\boldsymbol{x}=\sum_{i=1}^{p}x_i^2$ 服从自由度为 p，非中心参数为 $\lambda=\boldsymbol{\mu}'\boldsymbol{\mu}=\sum_{i=1}^{p}\mu_i^2$ 的非中心 χ^2 分布。记作 $y\sim\chi^2(p,\lambda)$。当 $\lambda=0$ 时，称 y 服从自由度为 p 的中心 χ^2 分布。记作 $y\sim\chi^2(p)$。

容易证明（留给读者）：

若 $y\sim\chi^2(p,\lambda)$ 则

1）$Ey=p+\lambda$　　(1.1)

2）$\operatorname{var}(y)=2p+4\lambda$　　(1.2)

下面我们计算 $y \sim \chi^2(p,\lambda)$ 的特征函数。

$$\varphi_\lambda(t) = E\exp(ity) = \prod_{j=1}^{p} E(\exp i x_j^2)$$

而

$$E\exp(itx_j^2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{itx_j^2 - \frac{1}{2}(x_j - \mu_j)^2\right\} dx_j$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(1-2it)\left(x_j - \frac{\mu_j}{1-2it}\right)^2 + \frac{it\mu_j^2}{1-2it}\right\} dx_j$$

$$= (1-2it)^{-\frac{1}{2}} \cdot \exp\{it\mu_j^2/(1-2it)\}$$

从而

$\chi^2(p,\lambda)$ 的特征函数为

$$\varphi_\lambda(t) = (1-2it)^{-\frac{p}{2}} \exp\left(\frac{it\lambda}{1-2it}\right) \tag{1.3}$$

当 $\lambda = 0$ 时 $\chi^2(p)$ 的特征函数为

$$\varphi_\lambda(t) = (1-2it)^{-p/2} \tag{1.4}$$

从一元统计分析中已知当 $y \sim \chi^2(p)$ 时，它的分布密度为

$$f(y;p) = \begin{cases} \left[\Gamma\left(\frac{p}{2}\right) \cdot 2^{\frac{p}{2}}\right]^{-1} y^{\frac{p}{2}-1} \exp\left(-\frac{1}{2}y\right) & y > 0 \\ 0 & \text{其他} \end{cases} \tag{1.5}$$

由于

$$\exp\left(\frac{it\lambda}{1-2it}\right) = e^{-\frac{\lambda}{2}} \exp\left(\frac{\lambda}{2(1-2it)}\right)$$

按幂级数展开得到 $\chi^2(p,\lambda)$ 的特征函数可展开成

$$\varphi_\lambda(t) = (1-2it)^{-\frac{p}{2}} \exp\left(\frac{it\lambda}{1-2it}\right)$$

$$= e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{\lambda}{2}\right)^j (1-2it)^{-(\frac{1}{2}p+j)}$$

根据特征函数的唯一性定理立即可得 $y \sim \chi^2(p,\lambda)$ 的密度为

$$f_\lambda(y;\ p) = \sum_{j=0}^{\infty} w_j(\lambda) \cdot f(y;p+2j) \tag{1.6}$$

其中 $w_j(\lambda) = \frac{1}{j!}\left(\frac{\lambda}{2}\right)^j e^{-\frac{1}{2}\lambda}$ 正是参数为 $\frac{\lambda}{2}$ 的 Poisson 概率分布的第 j 项概率，而 $f(y;\alpha)$ 正是 $\chi^2(\alpha)$ 的密度表达式

(2) t 分布

设 $x \sim N(\delta,1)$，$y \sim \chi^2(n)$ 且 x 与 y 相互独立，则称随机变量 $z = x\Big/\sqrt{\frac{1}{n}y}$ 服从自由度为 n，非中心参数为 $\lambda = \delta^2$ 的 t 分布，记为 $z \sim t(n,\lambda)$。当 $\lambda = 0$ 时称 z 服从自由度为 n 的中心 t 分布，记为 $z \sim t(n)$。

按一元统计知识知道当 $z \sim t(n)$ 时有密度

$$f(t;n) = \frac{n^{\frac{n}{2}}\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)}(n+t^2)^{-\frac{n+1}{2}} \tag{1.7}$$

当 $z \sim t(n,\lambda)$ 时有密度

$$f_\lambda(t;n) = e^{-\frac{\lambda}{2}}\frac{n^{n/2}}{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)(n+t^2)^{\frac{n+1}{2}}} \cdot \sum_{j=0}^{\infty}\Gamma\left(\frac{n+j+1}{2}\right)\frac{(\sqrt{\lambda}\,t)^j}{j!}\left(\frac{2}{n+t^2}\right)^{\frac{j}{2}} \tag{1.8}$$

(3) F 分布

当 $x \sim \chi^2(m,\lambda)$，$y \sim \chi^2(n,\nu)$ 且 x 与 y 相互独立，则称随机变量 $z = \frac{1}{m}x\Big/\frac{1}{n}y$ 服从自由度为 m 和 n，非中心参数为 λ 和 ν 的两重 F 分布，记为 $z \sim F(m,n,\lambda,\nu)$。而当 $\nu = 0$ 时称 z

服从自由度为 m 和 n，非中心参数为 λ 的 F 分布，记为 $z \sim F(m, n, \lambda)$。当 $\lambda = 0$，$\nu = 0$ 时，称 z 服从自由度为 m 和 n 的中心 F 分布，记为 $z \sim F(m,n)$ 今后我们仅讨论 $F(m,n,\lambda)$ 和 $F(m, n)$。

在(1.5)和(1.6)中已经给出了 $x \sim \chi^2(m,\lambda)$ 和 $y \sim \chi^2(n)$ 的分布密度，由此不难得到 $s = \frac{1}{m} x$ 和 $t = \frac{1}{n} y$ 的密度为 $g(s)$ 和 $h(t)$，那么 $z = s/t$ 的分布密度为

$$f_\lambda(z,m,n) = \int_0^\infty t g(zt) h(t) dt \quad (0 < z < \infty)$$

进行逐项积分，稍加整理即得 $z \sim F(m,n,\lambda)$ 的分布密度为

$$f_\lambda(z,m,n) = e^{-\frac{\lambda}{2}} \frac{n^{\frac{n}{2}} m^{\frac{m}{2}} z^{\frac{m}{2}-1}}{(n+mz)^{\frac{1}{2}(m+n)}} \sum_{i=0}^{\infty} \frac{c_i}{i!} \left(\frac{\lambda}{2}\right)^i \left(\frac{mz}{n+mz}\right)^i \quad (1.9)$$

其中

$$c_i = B\left(\frac{m}{2}+i, \frac{n}{2}\right)^{-1}$$

$$= \Gamma\left(\frac{m+n}{2}+i\right) \Big/ \left[\Gamma\left(\frac{m}{2}+i\right)\Gamma\left(\frac{n}{2}\right)\right]$$

当 $\lambda = 0$ 时 $z \sim F(m,n)$ 的分布密度为

$$f(z,m,n) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} n^{\frac{n}{2}} m^{\frac{m}{2}} \cdot \frac{z^{\frac{m}{2}-1}}{(n+mz)^{\frac{1}{2}(m+n)}}$$

(1.10)

下面是与 F 分布有密切关系的 β 分布。

(4) β 分布

若随机变量 x 的分布密度为

$$f(x) = \begin{cases} \dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1} & 0 < x < 1 \\ 0 & \text{其他} \end{cases}$$

则称 x 服从参数为 p, q 的 β 分布，记为 $x \sim \beta(p,q)$。

设 $x \sim \beta(p,q)$，则容易计算 x 的 h 阶矩为

$$Ex^h = \frac{\Gamma(p+q)\Gamma(p+h)}{\Gamma(p)\Gamma(p+q+h)} \tag{1.11}$$

且可以证明 β 的分布是由它的各阶矩唯一确定的。

下面几个性质是很有用的。

i) 如果 $u \sim t(n,\lambda)$，则 $u^2 \sim F(1,n,\lambda)$。

ii) 设 $x_1,\cdots,x_n$ 相互独立且都服从 $N(\mu,\sigma^2)$，则

$$z = \frac{\sqrt{n}(\bar{x}-a)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}} \sim t(n-1,\lambda) \tag{1.12}$$

其中 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\lambda = \frac{n(a-\mu)^2}{\sigma^2}$。特别当 $a=\mu$ 时，$z \sim t(n-1)$。

iii) 设 $x_1,\cdots,x_m; y_1,\cdots,y_n$ 全部相互独立，$x_i \sim N(\mu,\sigma^2)$ $(i=1,\cdots,m)$ $y_j \sim N(\nu,\sigma^2)(j=1,\cdots,n)$，则

$$z = \frac{\frac{1}{m-1}\sum_{i=1}^{m}(x_i-\bar{x})^2}{\frac{1}{n-1}\sum_{j=1}^{n}(y_j-\bar{y})^2} \sim F(m-1,n-1) \tag{1.13}$$

其中 $\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$，$\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j$。

iv) 设 $x \sim \chi^2(m)$，$y \sim \chi^2(n)$ 且相互独立，则

$$z_1 = \frac{x}{x+y} \sim \beta\left(\frac{m}{2}, \frac{n}{2}\right).$$

$$z_2 = \frac{y}{x+y} \sim \beta\left(\frac{n}{2}, \frac{m}{2}\right).$$

由于

$$u = \frac{1}{m} x \Big/ \frac{1}{n} y \sim F(m,n).$$

于是有

$$\frac{1}{1+\frac{m}{n} u} \sim \beta\left(\frac{n}{2}, \frac{m}{2}\right).$$

$$\frac{n}{m} \frac{1-z_2}{z_2} \sim F(m,n) \tag{1.14}$$

在下面各章讨论假设检验时经常用到上述各分布的分位点。以后将记 $\chi^2_\alpha(n)$表示 $\chi^2(n)$的下侧α分位点，即如 $z \sim \chi^2(n)$，则有 $P(z \leqslant \chi^2_\alpha(n)) = \alpha$。类似的有 $t_\alpha(n)$ 和 $F_\alpha(m,n)$，不再一一说明。

（二） 二次型的分布

设p维随机向量 $\boldsymbol{x} \sim N_p(\boldsymbol{0}, I_p)$. A是$p \times p$阶对称矩阵，首先我们考虑二次型 $y = \boldsymbol{x}' A \boldsymbol{x} = \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} x_i x_j$ 在什么条件下服从 χ^2 分布。为此我们先计算 $y = \boldsymbol{x}' A \boldsymbol{x}$ 的特征函数。

设 $\mathrm{rk}(A) = r$，按第一章定理 3.1，存在正交阵 P，使得

$$PAP' = \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_r, 0, \cdots, 0)$$

其中 $\lambda_1, \cdots, \lambda_r$ 是A的非零特征根。令 $\boldsymbol{z} = P\boldsymbol{x}$，由第二章定理 2.2 知 $\boldsymbol{z}$ 仍服从 $N_p(\boldsymbol{0}, I_p)$。此时 y 可写成

$$y = \boldsymbol{z}' \Lambda \boldsymbol{z} = \sum_{j=1}^{r} \lambda_j z_j^2 \quad (\text{其中 } \boldsymbol{z} = (z_1, \cdots, z_p)')$$

因此y的特征函数为

$$\varphi_y(t) = E[\exp(ity)] = E\left[\exp\left(it \sum_{j=1}^{r} \lambda_j z_j^2\right)\right]$$

而

$$E[\exp(i\lambda_j \cdot t \cdot z_j^2)] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} z_j^2 + it\lambda_j z_j^2\right\} dz_j$$

$$= (1 - 2it\lambda_j)^{-\frac{1}{2}}$$

因此

$$\varphi_y(t) = \prod_{j=1}^{r} (1 - 2it\lambda_j)^{-\frac{1}{2}} = |I - 2itA|^{-\frac{1}{2}} \tag{1.15}$$

对此只须注意到

$$|I - 2itA| = |P'||I - 2it\Lambda||P| = \prod_{j=1}^{r} (1 - 2it\lambda_j)$$

定理 1.1 设 $x \sim N_p(0, I_p), y = x'Ax$，其中 A 对称阵．则 $y \sim \chi^2(s)$ 的充分必要条件是

$$A^2 = A \quad 和 \quad s = \text{rk}(A)$$

证明 要证明 $y \sim \chi^2(s)$，只须证明它的特征函数(1.15)与 $\chi^2(s)$ 的特征函数(1.4)对于所有实数 t 恒等．即对于所有实数 t 有

$$\begin{aligned}(1 - 2it)^{-\frac{s}{2}} &= \prod_{j=1}^{r} (1 - 2it\lambda_j)^{-\frac{1}{2}} \\ &= \prod_{j=1}^{r} [1 - \lambda_j + (1 - 2it)\lambda_j]^{-\frac{1}{2}} \end{aligned} \tag{1.16}$$

记 $m = 1 - 2it$，(1.16) 可改写为

$$m^s = \left(\prod_{j=1}^{r} \lambda_j\right) \cdot \prod_{j=1}^{r} \left(m + \frac{1 - \lambda_j}{\lambda_j}\right)$$

比较多项式的指数和系数．得到(1.16)相等的充分必要条件是

$$s = r = \text{rk}(A) \quad 和 \quad \lambda_j = 1 \quad (j = 1, \cdots, r)$$

后者即 $\Lambda = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$，从而

$$A = P'\Lambda P = P'\Lambda P \cdot P'\Lambda P = A^2$$

即 A 是对称幂等阵．定理得证． ▌

我们将不加证明给出下列定理(参考[15])．

定理 1.2 设 $x \sim N_p(\mu,\Sigma)$, $\Sigma>0$, $y=x'Ax$, A 为对称阵. 则 $y\sim \chi^2(s,\lambda)$ 的充分必要条件是

$$\text{i) } A=A\Sigma A \quad \text{ii) } s=\mathrm{rk}(A) \quad \text{iii) } \lambda=\mu'A\mu \qquad (1.17)$$

特别地，如果定理中 $\mu=0$, 只须令 $z=\Sigma^{-\frac{1}{2}}x$, 此时 $z\sim N_p(0,I_p)$ 且 $y=z'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}z$, 由定理 1.1 知道 $y\sim\chi^2(s)$ 的充分必要条件是

$$(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})=(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})^2 \quad 和 \quad s=\mathrm{rk}(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}).$$

由 $\Sigma>0$,知道(1.17)成立.

推论 1.1 设 $x\sim N_p(\mu,\Sigma)\Sigma>0$, 则 $(x-\mu)'\Sigma^{-1}(x-\mu)\sim\chi^2(p)$

证明留给读者.

下面给出两个二次型相互独立的充分必要条件，为此我们先证明下述引理.

引理 1.1 设 $A_1,\cdots,A_k$ 是对称阵, $A=\sum\limits_{i=1}^{k}A_i$ 且

1) $A,A_1\cdots,A_{k-1}$ 是幂等阵

2) $A_k=B\cdot B'$,

那么 $A_i\cdot A_j=0(i\neq j)$, 且 A_k 是幂等阵, $\sum\limits_{i=1}^{k}\mathrm{rk}(A_i)=\mathrm{rk}(A)$.

证明 令 $A_0=I-A$, 由 A 幂等对称,知 A_0 也幂等对称且 $I=A_0+A_1+\cdots+A_k$. 于是对 $A_j(j=0,1,\cdots,k-1)$ 有

$$A_j(A_0+A_1+\cdots+A_k)A_j=A_j$$

令 $P_j=A_j(A_0,A_1,\cdots A_{j-1},A_{j+1},\cdots,A_{k-1},B)$ 其中括号中少一项 A_j. 那么利用条件 1) 可得

$$P_j\cdot P_j'=A_j\left(\sum_{i=0}^{k}A_i-A_j\right)A_j=0 \quad (j=0,1,\cdots,k-1)$$

所以 $\mathrm{rk}(P_j)=\mathrm{rk}(P_jP_j')=0$, 从而 $P_j=0$

于是

$$A_jA_i = 0 \quad (i \neq j;\ i,j = 0,1,\cdots,k-1)$$

且
$$A_jB = 0 \quad (j = 0,1,\cdots,k-1)$$

从而

$$A_jA_k = A_jBB' = 0 \quad (j = 0,1,\cdots,k-1)$$

$$A_k = \left(\sum_{i=0}^{k} A_i\right) A_k = A_k^2$$

即 A_k 是幂等阵，而幂等阵 A 有 $\mathrm{rk}A = \mathrm{tr}A$，立即可知

$$\sum_{i=1}^{k} \mathrm{rk}(A_i) = \mathrm{rk}(A). \quad \blacksquare$$

定理 1.3 设 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, I_p)$，$A_1, A_2$ 是 $p \times p$ 阶幂等对称阵，则 $\boldsymbol{x}'A_1\boldsymbol{x}$ 和 $\boldsymbol{x}'A_2\boldsymbol{x}$ 相互独立的充分必要条件是

$$A_1 \cdot A_2 = 0$$

证明 必要性：

按定理 1.2 知 $\boldsymbol{x}'A_i\boldsymbol{x} \sim \chi^2(r_i, \lambda_i)$，其中 $r_i = \mathrm{rk}(A_i)$，$\lambda_i = \boldsymbol{\mu}'A_i\boldsymbol{\mu}$ $i = 1, 2$. 令 $A = A_1 + A_2$，由于 $\boldsymbol{x}'A_1\boldsymbol{x}$ 和 $\boldsymbol{x}'A_2\boldsymbol{x}$ 相互独立。从它们的特征函数 (1.3) 的乘积立即可知 $\boldsymbol{x}'A\boldsymbol{x} \sim \chi^2(r_1 + r_2,\ \lambda_1 + \lambda_2)$。仍由定理 1.2 知 A 是幂等对称阵。于是由引理 1.1 知 $A_1A_2 = 0$.

充分性：

由 A_1, A_2 的幂等性知 $\boldsymbol{x}'A_i\boldsymbol{x} = (A_i\boldsymbol{x})'(A_i\boldsymbol{x}) \qquad (i = 1, 2)$
而 $A_i\boldsymbol{x}$ 是 $\boldsymbol{x}$ 的线性组合仍服从正态分布，且

$$\mathrm{cov}(A_1\boldsymbol{x}, A_2\boldsymbol{x}) = A_1\mathrm{cov}(\boldsymbol{x})A_2' = A_1 \cdot A_2 = 0$$

所以 $A_1\boldsymbol{x}$ 和 $A_2\boldsymbol{x}$ 相互独立。利用第一章定理 1.3 知道它们的函数 $\boldsymbol{x}'A_1\boldsymbol{x}$ 和 $\boldsymbol{x}'A_2\boldsymbol{x}$ 也相互独立。 $\blacksquare$

引理 1.2 设 $A_1, \cdots, A_k$ 是 $n \times n$ 阶对称阵，$\mathrm{rk}(A_i) = r_i$ $(i = 1, \cdots, k)$ $A = \sum_{i=1}^{k} A_i$ $\mathrm{rk}(A) = p$.

对于下列四个条件

1) A_i 都是幂等阵 $(i=1,\cdots,k)$

2) $A_i \cdot A_j = 0 \quad (i \neq j; i,j = 1,\cdots,k)$

3) A 幂等阵

4) $p = \sum_{i=1}^{k} r_i$

有以下结论

a) 1),2),3) 中任两个条件可推出其他两个。

b) 3),4) 条件可推出 1), 2).

证明见第三章附录。

定理 1.4 设 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \Sigma), \Sigma > 0, y_i = \boldsymbol{x}' A_i \boldsymbol{x} (i = 1, \cdots, k)$ 且 $y = \sum_{i=1}^{k} y_i = \boldsymbol{x}' A \boldsymbol{x}$. 那么对于下列四个条件

1) $y_i \sim \chi^2(r_i, \lambda_i)(i = 1, \cdots, k)$ 其中 $r_i = \mathrm{rk}(A_i)$, $\lambda_i = \boldsymbol{\mu}' A_i \boldsymbol{\mu}$

2) $y_1, \cdots, y_k$ 相互独立

3) $y \sim \chi^2(r, \lambda)$, 其中 $r = \sum_{i=1}^{k} r_i$, $\lambda = \sum_{i=1}^{k} \lambda_i$

4) $\mathrm{rk}(A) = \sum_{i=1}^{k} \mathrm{rk}(A_i)$.

有下列结论

a) 1),2),3)中任意两个条件可推出其他两个.

b) 3),4)二条件可推出 1), 2).

证明 由于 $A = \sum_{i=1}^{k} A_i$, 按定理 1.3 和定理 1.2 知道,上述四条件等价于

1)' $\Sigma^{\frac{1}{2}} A_i \Sigma^{\frac{1}{2}}$ 是幂等对称阵 $(i = 1, \cdots, k)$

2)' $(\Sigma^{\frac{1}{2}} A_i \Sigma^{\frac{1}{2}})(\Sigma^{\frac{1}{2}} A_j \Sigma^{\frac{1}{2}}) = 0$

3)' $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ 是幂等对称阵

4)′ $\mathrm{rk}(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})=\sum_{i=1}^{k}\mathrm{rk}(\Sigma^{\frac{1}{2}}A_i\Sigma^{\frac{1}{2}})$.

利用引理 1.2,立即得到所要的结论. ∎

定理 1.5 设 $x\sim N_p(\mu,\Sigma)$,$\Sigma>0$,$y_i=x'A_ix(i=1,\cdots,k)$,又 $y=\sum_{i=1}^{k}y_i=x'Ax$. 如果

1) $y,y_1,\cdots,y_{k-1}$ 都服从非中心 χ^2 分布

2) 对于所有 x,$y_k\geqslant 0$,

那么 $y_1,\cdots,y_k$ 相互独立且服从非中心 χ^2 分布.

证明 按定理 1.2 和非负定矩阵的性质知道条件 1), 2)等价于

1)′ $\Sigma^{\frac{1}{2}}A_i\Sigma^{\frac{1}{2}}(i=1,\cdots,k-1)$ 和 $\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}$ 是幂等对称阵.

2)′ $\Sigma^{\frac{1}{2}}A_k\Sigma^{\frac{1}{2}}=B\cdot B'$.

从而由引理 1.1 可得出结论. ∎

定理 1.4 的重要特例是有名的 Cochran 定理

推论 1.2 (Cochran 定理)

设 $x\sim N_p(\mu,I_p)$,$y=x'x=\sum_{i=1}^{k}y_i$,其中 $y_i=x'A_ix(A_i$ 是对称阵,$i=1,\cdots,k)$. 那么 $y_1,\cdots,y_k$ 相互独立且服从非中心 χ^2 分布的充分必要条件是

$$\sum_{i=1}^{n}\mathrm{rk}(A)=p.$$

证明 根据定理 1.4 知道条件 1), 2)和 3), 4) 等价. 因此 $y=x'x\sim\chi^2(p,\lambda)$,其中 $\lambda=\mu'\mu$ 以及 $\sum_{i=1}^{k}\mathrm{rk}(A_i)=\mathrm{rk}(I_p)=p$ 保证了 y_i 相互独立且分别服从 $\chi^2(r_i,\lambda_i)$,$r_i=\mathrm{rk}(A_i)$,$\lambda_i=\mu'A_i\mu$. 反之,若 $y_1,\cdots,y_k$ 相互独立且服从非中心 χ^2 分布则保证

$$\sum_{i=1}^{k} \mathrm{rk}(A_i) = p. \quad \blacksquare$$

§3.2 Wishart 分布及其性质

Wishart 分布是与多元正态分布有着密切联系的重要的多元分布，它是一元统计分析中 χ^2 分布的自然推广。对于实际工作者而言可以略去 Wishart 分布密度的推导过程，但关于它的定义及其有关性质仍要求掌握。

定义 2.1 设 $y_1,\cdots,y_n$ 相互独立且服从 p 维正态分布 $N_p(\mu_i,\Sigma)(i=1,\cdots,n)$，$\Sigma>0$ 且 $n\geqslant p$，即 $Y=(y_1,\cdots,y_n)\sim N_{p,n}(M,\Sigma,I_n)$，$M=(\mu_1,\cdots,\mu_n)$。那么我们称下列对称随机矩阵

$$A = YY' = \sum_{i=1}^{n} y_i y_i' = (a_{ij})_{p\times p}$$

服从的分布是一个自由度为 n 和 p，非中心参数为 $\Omega=MM'$ 的非中心 Wishart 分布，记为 $A\sim W_p(n,\Sigma,\Omega)$。当 $\Omega=0$ 时(即 $y_i\sim N_p(0,\Sigma)$)。称 A 服从的分布是自由度为 n 和 p 的中心 Wishart 分布。记为 $A\sim W_p(n,\Sigma)$。

类似于多元正态分布中的 $\Sigma\geqslant 0$，这里当 Σ 不是正定阵或者 $n<p$ 时，我们称 A 服从的分布是奇异的 Wishart 分布。

从定义 2.1 中我们注意到，虽然 Y 有 $p\times n$ 维变数，但 A 是对称的 $p\times p$ 阶矩阵，它实际上只有 $\frac{p(p+1)}{2}$ 维变数，因此 A 的分布实质上就是变数 $a_{11},a_{12},\cdots,a_{1p},a_{22},\cdots,a_{2p},\cdots,a_{pp}$ 的联合分布，而当 $p=1$ 且 $\sigma_{11}=1$ 时，A 正是通常的非中心 χ^2 分布。

下面我们将讨论当 $n\geqslant p$ 和 $\Sigma>0$ 条件下 Wishart 分布的分布密度。首先我们将证明，在 $n\geqslant p$ 及 $\Sigma>0$ 条件下 $A=YY'$ 以概率 1 正定，事实上可证明比这更广泛的结论。

引理 2.1 设 $y_i(i=1,\cdots,n,n\geqslant p)$ 是相互独立且诸 y_i 服

从非退化型连续型分布，令 $Y=(y_1,\cdots,y_n)$. 则 $A=YY'=\sum_{i=1}^{n}y_iy_i'$ 以概率 1 正定。

证明　考虑 $y_1,\cdots,y_n$ 中任意 p 个观察，则由

$$
\begin{aligned}
&P(\mathrm{rk}(y_1,\cdots y_p)=p)\\
&=1-P(y_1\cdots y_p \text{ 中至少有一个是其余 } p-1 \text{ 个的线性组合})\\
&\geqslant 1-\sum_{i=1}^{p}P\left(y_i=\sum_{j\neq i}a_{ij}y_j,\ \text{其中 } a_{ij} \text{ 是常数}\right)
\end{aligned}
$$

由于 y_i 相互独立且服从非退化连续型分布，所以对于 $i=1,\cdots,p$ 有

$$
P\left(y_i=\sum_{j\neq i}a_{ij}y_j \,\middle|\, y_1,\cdots y_{i-1},y_{i+1},\cdots,y_p\right)=0
$$

通过取条件期望，知道 $P(\mathrm{rk}(y_1\cdots y_p)=p)=1$. 而由

$$
\begin{aligned}
1&\geqslant P(\mathrm{rk}(A)=p)=P(\mathrm{rk}(Y)=p)\\
&\geqslant P(\mathrm{rk}(y_1,\cdots y_p)=p)=1.
\end{aligned}
$$

引理得证. ▮

引理 2.2　设 $y=(y_1,\cdots y_n)'\sim N_n(0,I_n)$，$Z$ 是 $r\times n$ 阶常数矩阵且 $\mathrm{rk}(Z)=r<n$，则

$$
x=Zy
$$

和

$$
A=y'y-x'(ZZ')^{-1}x=y'[I_n-Z'(ZZ')^{-1}Z]y
$$

相互独立且分别服从 $N_r(0,ZZ')$ 和 $\chi^2(n-r)$ 分布。

证明　记 $P=I_n-Z'(ZZ')^{-1}Z$ 易知它是幂等对称阵，且 $\mathrm{rk}(P)=n-r$，设 $t=Py$，于是 $A=y'y=t't$.

由第二章定理 2.2 知 $x=Zy\sim N_r(0,ZZ')$，由本章定理 1.1 知 $A\sim\chi^2(n-r)$. 另外

$$
\mathrm{cov}(x,t)=\mathrm{cov}(Zy,Py)=Z\cdot P=0
$$

而 x,t 服从正态分布，所以 x,t 相互独立. 由第二章定理 1.3 知它们的函数 x 和 A 也相互独立. ▮

定理 2.1 设 $Z=(z_1,\cdots,z_n)\sim N_{p,n}(0,I_p,I_n)(n\geqslant p)$. 那么

$$A=ZZ'=\sum_{i=1}^{n}z_iz_i'$$

的分布密度是由下式给出的:

$$f(A)=\begin{cases}c(p,n)|A|^{\frac{1}{2}(n-p-1)}\operatorname{etr}\left(-\frac{1}{2}A\right) & A>0\\ 0 & \text{其他}\end{cases}\tag{2.1}$$

其中 $c(p,n)=\left[\pi^{\frac{1}{4}p(p-1)}\prod_{i=1}^{p}\Gamma\left(\frac{n-i+1}{2}\right)\cdot 2^{\frac{1}{2}np}\right]^{-1}$

*)证明 用归纳法. 当 $p=1$ 时 $A\sim\chi^2(n)$, 它的密度正是(2.1)式中令 $p=1$ 所得到的,即上节(1.5)式. 设当维数为 $p-1$ 时定理成立. 当维数为 p 时,作如下分块:

$$Z=\begin{pmatrix}z^{(1)'}\\ Z_2\end{pmatrix},\quad A=\begin{pmatrix}a_{11} & a_1'\\ a_1 & A_{22}\end{pmatrix},$$

其中 $z^{(1)}$ 是 n 维向量, Z_2 是 $(p-1)\times n$ 阶随机矩阵 A_{22} 是 $(p-1)\times(p-1)$ 阶负定矩阵,且

$$a_{11}=z^{(1)'}z^{(1)},\quad a_1=Z_2z^{(1)},\quad A_{22}=Z_2Z_2'$$

$$a_{1\cdot 2}=a_{11}-a_1A_{22}^{-1}a_1=z^{(1)}[I-Z_2'(Z_2Z_2')^{-1}Z_2]z^{(1)'}$$

由于

$$f(A)=f(a_{11},a_1,A_{22})=f(a_{11},a_1|A_{22})\cdot f(A_{22})$$

其中 $f(a_{11},a_1|A_{22})$ 是在 A_{22} 给定条件下 a_{11}, a_1 的条件分布密度. 由于在 A_{22} 给定下 $J(a_{11},a_1\to a_{1\cdot 2},a_1)=1$, 故

$$f(A)=f(a_{1\cdot 2},a_1|A_{22})f(A_{22}).\tag{2.2}$$

由归纳法知, $Z_2\sim N_{p-1,n}(0,I_{p-1},I_n)$ 和

$$f(A_{22})=c(p-1,n)|A_{22}|^{\frac{1}{2}(n-p)}\operatorname{etr}\left(-\frac{1}{2}A_{22}\right)\tag{2.3}$$

而当 Z_2 给定时,由引理 2.2 知道

$$a_1=Z_2z^{(1)}\sim N_{p-1}(0,Z_2Z_2')$$

$$a_{1\cdot 2} \sim \chi^2(n-p+1)$$

且相互独立，由于 $\boldsymbol{a}_1$ 的分布仅通过 $A_{22}=Z_2Z_2'$ 依赖于 Z_2，从而

$$f(a_{1\cdot 2},\boldsymbol{a}_1|A_{22})=(2\pi)^{-\frac{p-1}{2}}|A_{22}|^{-\frac{1}{2}}\cdot \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{a}_1'A_{22}^{-1}\boldsymbol{a}_1\right)$$

$$\cdot a_{1\cdot 2}^{\frac{1}{2}(n-p)}\cdot \operatorname{etr}\left(-\frac{1}{2}a_{1\cdot 2}\right)\left[2^{\frac{n-p+1}{2}}\right.$$

$$\left.\cdot \Gamma\left(\frac{n-p+1}{2}\right)\right]^{-1} \tag{2.4}$$

将(2.3),(2.4)代入(2.2)式中,利用

$$|A|=|A_{22}|\cdot a_{1\cdot 2},\quad \operatorname{tr}(A_{22}+a_{1\cdot 2}+\boldsymbol{a}_1'A_{22}^{-1}\boldsymbol{a}_1)=\operatorname{tr}A$$

再比较 e 的指数和 2 的指数，即可计算得到 A 的分布密度正是(2.1)式。 ▌

定理 2.2 设 $Y=(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n)\sim N_{p,n}(0,\Sigma,I_n)$ $(\Sigma>0, n\geqslant p)$ 那么 $V=Y\cdot Y'$ 的密度(即 $V\sim W_p(n,\Sigma)$ 的密度)为

$$f(V)=\begin{cases}c(p,n)\cdot|\Sigma|^{-\frac{n}{2}}|V|^{\frac{1}{2}(n-p-1)}\operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}V\right) & V>0\\ 0 & \text{其他}\end{cases} \tag{2.5}$$

证明 由于 Σ 正定，存在 $\Sigma^{\frac{1}{2}}>0$。令 $Z=\Sigma^{-\frac{1}{2}}Y$，按第二章定理 2.1 知，$Z\sim N_{p,n}(0,I_p,I_n)$，记 $A=ZZ'$，由定理 2.1 知 A 有密度(2.1).

在 (2.1) 式作变换 $A=\Sigma^{-\frac{1}{2}}V\Sigma^{-\frac{1}{2}}$，按第一章定理 4.3，$J(A\to V)=|\Sigma|^{-\frac{p+1}{2}}$ 并注意到 $\operatorname{tr}A=\operatorname{tr}\Sigma^{-\frac{1}{2}}V\Sigma^{-\frac{1}{2}}=\operatorname{tr}\Sigma^{-1}V$ 和 $|A|=|\Sigma^{-1}||V|$ 即可证明 V 的密度是由(2.5)式给出的。 ▌

关于非中心 Wishart 分布的分布密度的讨论已超出本书的范围。但有关它们的性质我们仍要进行讨论。

从(2.5)式可以知道 $W_1(n,\sigma^2)$ 正是 $\sigma^2\cdot\chi^2(n)$。

下面我们从定理 2.1 导出一个十分有用的推论。

[*]推论 2.1 令 Y 是 $p\times n$ 阶随机矩阵（$n\geqslant p$ 且 $\mathrm{rk}(Y)=p$），那么

$$\int_{YY'=A}dY=\frac{\pi^{\frac{1}{2}pn}}{\Gamma_p\left(\frac{1}{2}n\right)}|A|^{\frac{1}{2}(n-p-1)}\quad (A>0)\quad (2.6)^{(*)}$$

其中 $\Gamma_p\left(\frac{n}{2}\right)=\pi^{\frac{1}{4}p(p-1)}\cdot\sum_{i=1}^{p}\Gamma\left(\frac{n-i+1}{2}\right)$

证明 设 Y 是 $p\times n$ 阶随机矩阵，且 $Y\sim N_{p,n}(0,I_p,I_n)$ 那么由定理 2.1 知 $A=YY'$ 有密度(2.1)式。另一方面 A 的密度又可以如下求得：

$$f(A)=\int_{YY'=A}f(Y)dY$$

其中 $f(Y)$ 是 Y 的分布密度，从而

$$\begin{aligned}f(A)&=\int_{YY'=A}(2\pi)^{-\frac{np}{2}}\mathrm{etr}\left(-\frac{1}{2}YY'\right)dY\\&=(2\pi)^{-\frac{np}{2}}\mathrm{etr}\left(-\frac{1}{2}A\right)\cdot\int_{YY'=A}dY\\&=c(p,n)|A|^{\frac{n-p-1}{2}}\mathrm{etr}\left(-\frac{1}{2}A\right)\end{aligned}$$

由于 A 以概率 1 正定，$|A|>0$，所以

$$\begin{aligned}\int_{YY'=A}dY&=c(p,n)|A|^{\frac{n-p-1}{2}}(2\pi)^{\frac{np}{2}}\\&=\frac{\pi^{\frac{1}{2}pn}}{\Gamma_p\left(\frac{n}{2}\right)}|A|^{\frac{n-p-1}{2}}\quad (A>0).\end{aligned}$$ ∎

利用(2.6)式我们可得到一些重要的公式。

例 2.1 求 n 维球体 $x'x=R^2$ 的体积和它的表面积

解 只须在(2.6)式中令 $p=1$，于是

(*) 这里 $\int_{YY'=A}dY$ 应理解为在作一对一变换 $Y\to(A,T)$ 后，固定 A 在 T 的值域上进行积分。

$$\int_{x'x=r^2} dx = \frac{\pi^{\frac{1}{2}n}}{\Gamma\left(\frac{n}{2}\right)} (r^2)^{\frac{n-2}{2}}$$

从而

$$V_R = \int_0^{R^2} \left[\int_{x'x=r^2} dx\right] dr^2 = \int_0^{R^2} \frac{\pi^{\frac{1}{2}n}}{\Gamma\left(\frac{n}{2}\right)} (r^2)^{\frac{n-2}{2}} dr^2$$

$$= \frac{2\pi^{\frac{1}{2}n}}{n\Gamma\left(\frac{n}{2}\right)} R^n \tag{2.7}$$

当 $n=2$ 时即为圆面积，$n=3$ 时利用 $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$ 和 $\Gamma(p+1)=p\Gamma(p)$ 即可求得球体积为 $\frac{4}{3}\pi R^3$。

设 $x'x=r^2$ 的球面面积为 S_r，那么

$$V_R = \int_0^R S_r dr.$$

将(2.7)中 dr^2 改为 $2rdr$ 立即可知

$$S_r = \frac{2\pi^{\frac{1}{2}n}}{\Gamma\left(\frac{n}{2}\right)} r^{n-1}.$$

当 $n=2$ 时即为圆周长 $2\pi r$。$n=3$ 时即为球面表面积 $4\pi r^2$。

例 2.2　利用(2.6)式求出 $\chi^2(n)$ 的分布密度

解　设 $t=x'x$ 其中 $x=(x_1,\cdots,x_n)'\sim N_n(0,I_n)$，则 $t\sim\chi^2(n)$。于是它的密度为

$$g(t) = \int_{x'x=t} f(x)dx = \int_{x'x=t} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}x'x} dx$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} t^{\frac{n}{2}-1} e^{-\frac{1}{2}t} \qquad (t>0)$$

推论 2.2　设 $p\times n$ 阶随机矩阵 Y 的密度为 $f(YY')$，又设

$V = YY'$. 那么V的密度为

$$g(V) = \frac{\pi^{\frac{1}{2}pn}}{\Gamma_p\left(\frac{n}{2}\right)} |V|^{\frac{n-p-1}{2}} f(V) \quad (V > 0)$$

证明　直接利用推论 2.1 即可证明. ▮

推论 2.2 是一个很有用的工具,它指出只要 Y 的密度是 $f(Y \cdot Y')$ 形式,而不必限于正态分布,那么 $V = YY'$ 的分布总可以求得. 它可以帮助我们解决多元统计分析中除了正态分布以外其他分布的许多问题. 近年来国际上许多统计学家和我国一些学者曾对椭球等高分布族进行过深入的研究，得到与正态分布相类似的良好结果.

一个 p 维随机向量 $\boldsymbol{x}$，如果它的密度存在且有形式

$$|\Sigma|^{-\frac{1}{2}} g[(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})]$$

其中 $\Sigma > 0$，那么我们称 $\boldsymbol{x}$ 服从 p 维椭球等高分布,当 $g(t) = c \cdot e^{-\frac{1}{2}t}$ 时正是正态分布. 当 $g(t)$ 取其他形式时还可得到一些常见的多元分布，对于这一类分布的平方和和其他有关的问题都可以利用推论 2.2 来处理，感兴趣的读者可参考文献(22)的 VIII.

下面我们研究 Wishart 分布的若干性质,其中许多是 χ^2 分布性质的推广.

定理 2.3　设 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 相互独立且都服从 $N_p(\boldsymbol{\theta}, \Sigma)$, $(\Sigma > 0, n > p)$. 即 $X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) \sim N_{p,n}(\boldsymbol{\theta} \cdot \mathbf{1}_n', \Sigma, I_n)$, 令

$$\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i = \frac{1}{n} X \cdot \mathbf{1}_n$$

$$\begin{aligned} A &= \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})' \\ &= \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' - n\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}' = XX' - n\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}' \\ &= X\left[I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right]X' \end{aligned}$$

那么

$$\bar{x} \sim N_p\left(\theta, \frac{1}{n}\Sigma\right), \quad A \sim W_p(n-1, \Sigma) \text{ 且相互独立.}$$

证明　令 P 是 $n \times n$ 阶正交矩阵，它的第一列为 $\frac{1}{\sqrt{n}}\mathbf{1}_n$，那么

$$\mathbf{1}_n' P = (\sqrt{n}, 0, \cdots, 0).$$

由第二章定理 4.1 有

$$Z = (X - \theta \mathbf{1}_n')P = (\sqrt{n}(\bar{x} - \theta), Y) \sim N_{p,n}(0, \Sigma, I_n)$$

其中 Y 是 $p \times (n-1)$阶矩阵且与θ无关，而$J(X \to \sqrt{n}\,\bar{x}, Y) = |PP'|^{\frac{p}{2}} = 1$ 作变换 $\dot{X} \to Z$ 并注意到

$$(X - \theta \mathbf{1}_n')(X - \theta \mathbf{1}_n')' = n(\bar{x} - \theta)(\bar{x} - \theta)' + YY'$$

从X的密度表达式可求出$\sqrt{n}\,\bar{x}$ 和Y的联合分布密度为

$$(2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}\operatorname{etr}\left\{-\frac{1}{2}\Sigma^{-1}[(\sqrt{n}\,\bar{x} - \sqrt{n}\theta)\right.$$

$$\left.\cdot(\sqrt{n}\,\bar{x} - \sqrt{n}\theta)' + YY']\right\}$$

由此可知

$$\sqrt{n}\,\bar{x} \sim N_p(\sqrt{n}\theta, \Sigma),$$

$$y \sim N_{p,n-1}(0, \Sigma, I_{n-1})$$

且相互独立．而

$$A = XX' - n\bar{x} \cdot \bar{x}' = (X - \theta \mathbf{1}_n')(X - \theta \mathbf{1}_n')$$

$$n(\bar{x} - \theta)(\bar{x} - \theta)' = YY'$$

按定理 2.1 知 $A \sim W_p(n-1, \Sigma)$。由第二章定理 1.3 知 $\bar{x}$ 与A独立。而 $\bar{x} \sim N_p\left(\theta, \frac{1}{n}\Sigma\right)$ 是显然的。 ▮

注意到 $S = \frac{1}{n-1}A$ 正是通常样本协方差矩阵，又由 $n >$

p 所以 $n-1 \geqslant p$ 保证了 A 以概率 1 正定. 而 $\bar{x}$ 正是通常样本均值向量.

下面给出当 $A \sim W_p(n, \Sigma, \Omega)$ 时 A 的相应子矩阵的边缘分布.

定理 2.4 设 $A \sim W_p(n, \Sigma, \Omega)$ 且 A 有划分

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

其中 A_{11} 是 q 阶方阵 $(1 \leqslant q \leqslant p)$ 而 Σ_{ij}, Ω_{ij} 为相应分块矩阵, 那么

$$A_{11} \sim W_q(n, \Sigma_{11}, \Omega_{11}).$$

证明 由于 $A \sim W_p(n, \Sigma, \Omega)$, 按定义存在 $p \times n$ 阶随机矩阵 $Y \sim N_{p,n}(M, \Sigma, I_n)$ 使得 $A = Y \cdot Y'$, 其中 $\Omega = MM'$.

记 $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, 其中 Y_1 为 $q \times n$ 阶矩阵, 那么按第二章推论 2.1 知 $Y_1 \sim N_{q,n}(M_1, \Sigma_{11}, I_n)$, 其中 $M = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix}$, M_1 是 $q \times n$ 阶矩阵

由于 $A_{11} = Y_1 \cdot Y_1'$, 从而 $A_{11} \sim W_q(n, \Sigma_{11}, \Omega_{11}), \Omega_{11} = M_1 M_1'$. ▉

定理 2.5 设 $A \sim W_p(n, \Sigma, \Omega)$, $A = (A_{ij}), \Omega = (\Omega_{ij})$, 其中 A_{ij} 为 $q_i \times q_j$ 阶矩阵 $(i, j = 1, \cdots, k)$ $\sum_{i=1}^{k} q_i = p, \Sigma_{ij}, \Omega_{ij}$ 是相应的分块矩阵, 则

$$A_{ii} \sim W_{q_i}(n, \Sigma_{ii}, \Omega_{ii})$$

证明 证法与定理 2.4 相仿, 只是令 $A = Y \cdot Y'$, 其中 $Y \sim N_{p,n}(M, \Sigma, I_n)$, 以及 $Y = (Y_1', Y_2', \cdots Y_k')'$, $Y_i \sim N_{q_i,n}(M_i, \Sigma_{ii}, I_n)$, 所以

$$A_{ii} = Y_i Y_i' \sim W_{q_i}(n, \Sigma_{ii}, \Omega_{ii}), \Omega_{ii} = M_i M_i'. \quad ▉$$

推论 2.3 设 $A \sim W_p(n, \Sigma, \Omega)$, 其中 $\Sigma = (\sigma_{ij}), \Omega = (\omega_{ij})$, $A = (a_{ij})$ 都是 $p \times p$ 阶矩阵, 那么

$$a_{ii} \sim \sigma_{ii}\chi^2\left(n, \frac{\omega_{ii}}{\sigma_{ii}}\right).$$

如果 $\Sigma = I_p$，那末 a_{ii} 相互独立且服从 $\chi^2(n, \omega_{ii})$。

证明留给读者。

定理 2.6 设 $A \sim W_p(n, \Sigma)$ $n \geqslant p$，$\Sigma > 0$，以及

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

其中 A_{11} 是 $r \times r$ 阶矩阵，A_{22} 是 $s \times s$ 阶矩阵，$r + s = p, \Sigma_{ij}$ 为相应的分块，并记 $A_{1\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$。 $\Sigma_{1\cdot 2} = \Sigma_{11} - \Sigma_{12} \cdot \Sigma_{22}^{-1}\Sigma_{21}, \beta = \Sigma_{12}\Sigma_{22}^{-1}$，那末

1）$A_{1\cdot 2} \sim W_r(n - s, \Sigma_{1\cdot 2})$ 且与 (A_{12}, A_{22}) 相互独立。

2）给定 A_{22} 后 $A_{12} \sim N_{r,s}(\beta A_{22}, \Sigma_{1\cdot 2}, A_{22})$

3）$A_{22} \sim W_s(n, \Sigma_{22})$

*)证明 在 A 的密度表达式 (2.5) 中作变换 A_{11}，A_{12}，$A_{22} \to A_{1\cdot 2}, A_{12}, A_{22}$，由第一章定理 4.1 知 $J = J(A_{11}, A_{12}, A_{22} \to A_{1\cdot 2}, A_{12}, A_{22}) = 1$，利用

$$A = \begin{pmatrix} I & A_{12}A_{22}^{-1} \\ 0 & I \end{pmatrix}\begin{pmatrix} A_{1\cdot 2} & 0 \\ 0 & A_{22} \end{pmatrix}\begin{pmatrix} I & 0 \\ A_{22}^{-1}A_{21} & I \end{pmatrix}$$

和

$$\Sigma^{-1} = \begin{pmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix}\begin{pmatrix} \Sigma_{1\cdot 2}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix}\begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix}$$

经过简单计算可得

$$\begin{aligned} \operatorname{tr}\Sigma^{-1}A = \operatorname{tr} & \begin{pmatrix} \Sigma_{1\cdot 2}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix}\begin{pmatrix} I & -\beta \\ 0 & I \end{pmatrix}\begin{pmatrix} I & A_{12}A_{22}^{-1} \\ 0 & I \end{pmatrix} \\ & \cdot \begin{pmatrix} A_{1\cdot 2} & 0 \\ 0 & A_{22} \end{pmatrix}\begin{pmatrix} I & 0 \\ A_{22}^{-1}A_{21} & I \end{pmatrix}\begin{pmatrix} I & 0 \\ -\beta & I \end{pmatrix} \\ = \operatorname{tr} & [\Sigma_{1\cdot 2}^{-1}A_{1\cdot 2} + \Sigma_{1\cdot 2}^{-1}(A_{12} - \beta A_{22})A_{22}^{-1}(A_{12} \\ & - \beta A_{22})' + \Sigma_{22}^{-1}A_{22}] \end{aligned}$$

代入(2.5)式之中并乘以 J 得到 $A_{1\cdot 2}, A_{12}, A_{22}$ 的联合密度为

$$f(A_{1\cdot 2}, A_{12}, A_{22}) = f_1(A_{1\cdot 2}) f_2(A_{12} | A_{22}) \cdot f_3(A_{22})$$

其中

$$f_1(A_{1\cdot 2}) = c_1 |\Sigma_{1\cdot 2}|^{-\frac{n-s}{2}} |A_{1\cdot 2}|^{\frac{n-s-r-1}{2}} \text{etr} \cdot \left(-\frac{1}{2} \Sigma_{1\cdot 2}^{-1} A_{1\cdot 2}\right)$$

$$f_2(A_{12} | A_{22}) = c_2 |\Sigma_{1\cdot 2}|^{-\frac{s}{2}} |A_{22}|^{-\frac{r}{2}} \text{etr} \cdot \left[-\frac{1}{2} \Sigma_{1\cdot 2}^{-1} (A_{12} - \beta A_{22})' A_{22}^{-1} (A_{12} - \beta A_{22})\right]$$

$$f_3(A_{22}) = c_3 |\Sigma_{22}|^{-\frac{n}{2}} |A_{22}|^{\frac{n-s-1}{2}} \text{etr} \left(-\frac{1}{2} \Sigma_{22}^{-1} A_{22}\right)$$

这里各 c_i $(i=1,2,3)$ 分别是规范化常数，使得密度在定义域上积分为 1。由因子分解定理知道

$$A_{1\cdot 2} \sim W_r(n-s, \Sigma_{1\cdot 2}) \text{ 且与 } (A_{12}, A_{22}) \text{ 独立.}$$

$$A_{12} | A_{22} \sim N_{r,s}(\beta A_{22}, \Sigma_{1\cdot 2}, A_{22})$$

$$A_{22} \sim W_s(n, \Sigma_{22}).$$

定理得证。 ▮

定理 2.7 设 $A \sim W_p(n, \Sigma, \Omega)$ $\Sigma > 0, n \geq p$, B 为 $m \times p$ 阶矩阵且 $\text{rk}(B) = m$，则

$$BAB' \sim W_m(n, B\Sigma B', B\Omega B').$$

证明 由于 $A \sim W_p(n, \Sigma, \Omega)$。存在 $Y \sim N_{p,n}(M, \Sigma, I_n)$，$\Omega = M \cdot M'$ 使得 $A = YY'$。而

$$BY \sim N_{m,n}(BM, B\Sigma B', I_n)$$

所以按定义有

$$BAB' = BY \cdot (BY)' \sim W_m(n, B\Sigma B', B\Omega B'). \quad ▮$$

推论 2.4 设 $A \sim W_p(n, \Sigma, \Omega)$，则 A 的任何二次型

$$l'Al \sim \sigma_l^2 \chi^2(n, \lambda)$$

其中 l 为 p 维向量，$\sigma_l^2 = l'\Sigma l$，$\lambda = l'\Omega l / \sigma_l^2$。

而当 $A \sim W_p(n, \Sigma)$ 时，$l'Al \sim \sigma_l^2 \chi^2(n)$。

证明留给读者。

定理 2.8 设 $X \sim N_{p,k}(M, \Sigma, I_k)(\Sigma > 0)$，$l$ 是 p 维非零向量，B 是 $k \times k$ 阶对称矩阵，$\mathrm{rk}(B) = n$，$k \geqslant n \geqslant p$。那么

$$A = XBX' \sim W_p(n, \Sigma, \Omega)$$

的充分必要条件是

对于上述任一非零向量 l，有

$$l'Al \sim \sigma_l^2 \chi^2(n, \lambda)$$

其中 $\Omega = MBM'$，$\sigma_l^2 = l'\Sigma l$，$\lambda = l'\Omega l/\sigma_l^2$。

证明 必要性由定理 2.7 得到，若 $A = XBX' \sim W_p(n, \Sigma, \Omega)$，则 $l'Al \sim \sigma_l^2\chi^2(n, \lambda)$。

充分性：

由于 $X'l \sim N_k(M'l, \sigma_l^2 I_k)$。以及 $(l'X)B(X'l) \sim \sigma_l^2\chi^2(n, \lambda)$，由定理 1.2 知 B 是对称幂等阵，且 $\mathrm{rk}(B) = n$，于是由第一章定理 3.2 知道存在 $k \times n$ 矩阵，使得 $B = UU'$ 且 $U'U = I_n$。因此由第二章定理 4.1 知

$$XU \sim N_{p,n}(MU, \Sigma, I_n)$$

因此

$$XBX' = (XU) \cdot (XU)' \sim W_p(n, \Sigma, \Omega) \quad \text{其中 } \Omega = MBM'.$$

此外由于 $l'\Omega l = 0$ 对一切非零向量 l 成立等价于 $\Omega = 0$，所以 $A \sim W_p(n, \Sigma)$ 的充分必要条件是对任一非零向量 l 有 $l'(XBX')l \sim \sigma_l^2\chi^2(n), n = \mathrm{rk}(B)$。▮

推论 2.5 设 $X \sim N_{p,n}(M, \Sigma, I_n)$，$C$ 是 $n \times n$ 阶对称矩阵，则

1) $XCX' \sim W_p(r, \Sigma, \Omega)$ 的充分必要条件是

$$C^2 = C, r = \mathrm{rk}(C), \ \Omega = MCM'.$$

2) $XCX' \sim W_p(r, \Sigma)$ 的充分必要条件是

$$C^2 = C, \ r = \mathrm{rk}(C), \Omega C = 0.$$

证明 由定理 2.8，定理 1.2，立即可得。▮

定理 2.9 设 $X \sim N_{p,n}(M, \Sigma, I_n)(n \geqslant p, \Sigma > 0)$，$C_1$ 和 C_2

是两个 $n \times n$ 阶对称幂等矩阵，则 XC_1X' 和 XC_2X' 相互独立的充分必要条件是 $C_1C_2 = 0$。

证明　必要性：如果 XC_1X' 和 XC_2X' 相互独立，则由第二章定理 1.3 知对任一 p 维向量 l，$l'XC_1X'l$ 和 $l'XC_2X'l$ 也相互独立，由于 $X'l \sim N_n(M'l, \sigma_l^2 I_n)$，所以由定理 1.3 知 $C_1 \cdot C_2 = 0$。

充分性：若 $C_1C_2 = 0$，由第二章定理 4.2 知 XC_1 和 XC_2 相互独立，从而它们各自可测函数 $(XC_1)(XC_1)' = XC_1X'$ 和 $(XC_2)(XC_2)' = XC_2X'$ 按第二章定理 1.3 也相互独立. ▌

定理 2.10　设 $X \sim N_{p,n}(M, \Sigma, I_n)\ (n \geq p, \Sigma > 0)$，$C_1, \cdots, C_k$ 是 $n \times n$ 阶对称矩阵. $C = \sum_{i=1}^{k} C_i$　$r_i = \text{rk}(C_i)$　$r = \text{rk}(C)$，那么下列四个条件中

1) $\{C_i, i = 1, \cdots, k\}$ 是幂等阵

2) $C_i \cdot C_j = 0\ (i \neq j, i, j = 1, \cdots, k)$

3) C 是幂等阵

4) $r = \sum_{i=1}^{k} r_i$

当 1),2),3) 中有两个成立或者 3),4) 同时成立时都有如下的结论：

a) $XC_iX' \sim W_p(r_i, \Sigma, \Omega_i)$ 其中 $\Omega_i = MC_iM'(i = 1, \cdots, k)$。

b) $\{XC_iX', i = 1, \cdots, k\}$ 相互独立。

c) $XCX' \sim W_p(r, \Sigma, \Omega)$ 其中 $\Omega = MCM', r = \sum_{i=1}^{k} r_i$。

证明　由引理 1.2 知道上述四个条件当 1),2),3)中任两个或者 3),4)同时成立时，四个条件同时成立. 于是由推论 2.5 和定理 2.9 可推出 a),c),b)。 ▌

推论 2.6 （多元 Cochran 定理）

设 $X \sim N_{p,n}(M, I_p, I_n)$ $n \geqslant p$, $A = XX' = \sum_{i=1}^{k} A_i$ 其中 $A_i = XB_iX'$, B_i 是 $n \times n$ 阶对称矩阵 $i = 1, \cdots, k$. 那么 A_i 相互独立且服从非中心 Wishart 分布的充分必要条件是

$$\sum_{i=1}^{k} \operatorname{rk}(B_i) = n. \tag{2.8}$$

证明 在定理 2.10 中取 $C_i = B_i, C = I_n (i = 1, \cdots, k)$. 则由于 I_n 显然幂等的、由(2.8)式知条件 3), 4) 成立,从而推论成立,反之是显然的,留给读者证明. ▮

从上述诸定理可以看到 Wishart 分布的许多性质完全是 χ^2 分布的性质的多维推广，故在此不一一列出. 例如与定理 1.5 相应的性质读者可自行写出.

下面我们求出 $A \sim W_p(n, \Sigma)$ 的特征函数,注意到此时它对 n 和 Σ 不再有 $n \geqslant p$ 和 $\Sigma > 0$ 的限制，也就是说即使 A 是奇异的 Wishart 分布时也同样成立.

定理 2.11 设 $A \sim W_p(n, \Sigma)$，那么 A 的特征函数是

$$\varphi_A(T) = |I_p - 2i\Sigma T|^{-\frac{n}{2}} \tag{2.9}$$

在定理 2.11 证明前，我们先作如下说明：由于 A 中随机变数仅为上三角矩阵元素 $a_{11}, \cdots, a_{1p}, a_{22}, \cdots, a_{2p}, \cdots a_{pp}$，从而它的特征函数应该是 $E\left\{\exp\left(i \sum_{k=1}^{p} \sum_{j=k}^{p} a_{kj} t_{kj}\right)\right\}$，然而这样处理有所不便，我们将记 $T = (t_{kj})$ 是一 $p \times p$ 阶对称矩阵，以 $E[\operatorname{etr} \cdot (iT'A)] = E\left[\exp\left(i \sum_{k=1}^{p} \sum_{j=1}^{p} a_{kj} t_{kj}\right)\right]$ 作为 A 的特征函数,请记住此时 A 的特征函数是 $a_{11}, 2a_{12}, \cdots, 2a_{1p}, a_{22}, \cdots, 2a_{2p}, \cdots, a_{pp}$ 的特征函数。

*证明　设 $A \sim W_p(n, \Sigma)$，$\mathrm{rk}(\Sigma) = r$，按定义存在 $Y = (y_1 \cdots y_n) \sim N_{p,n}(0, \Sigma, I_n)$，使得 $A = Y \cdot Y'$。即诸 y_i 相互独立且服从 $\sim N_p(0, \Sigma)$，按第二章正态分布定义，可将 y_i 表示为 $y_i = Bz_j$，其中 z_i 相互独立且服从 $N_r(0, I_r)$，而 B 是 $p \times r$ 阶矩阵且 $\sim BB' = \Sigma$ 且 $\mathrm{rk}(B) = r$。于是 A 的特征函数可表示为：

对于 $p \times p$ 阶对称矩阵 $T = (t_{kj})$，

$$
\begin{aligned}
\varphi_A(T) &= E(\mathrm{etr}\, iT'A) \\
&= E[\mathrm{etr}(iY'TY)] \\
&= \prod_{j=1}^{n} E[\exp(iz_j'B'TBz_j]
\end{aligned}
$$

由本章(1.15)式知 $z_j'B'TBz_j$ 的特征函数是

$$|I_r - 2itB'TB|^{-\frac{1}{2}}$$

令 $t = 1$ 且利用第一章推论 2.3，有

$$|I_r - 2iB'TB| = |I_p - 2i\Sigma T|$$

因此

$$
\begin{aligned}
\varphi_A(T) &= \prod_{j=1}^{n} |I_r - 2iB'TB|^{-\frac{1}{2}} \\
&= |I_p - 2i\Sigma T|^{-\frac{n}{2}}. \quad \blacksquare
\end{aligned}
$$

下面不加证明给出非中心 Wishart 分布的特征函数。

定理 2.12　设 $A \sim W_p(n, \Sigma, \Omega)$，那么 A 的特征函数为

$$\varphi_A(T) = |I_p - 2i\Sigma T|^{-\frac{n}{2}} \mathrm{etr}[iT(I_p - 2i\Sigma T)^{-1}\Omega] \tag{2.10}$$

定理 2.13　设 $A_j \sim W_p(n_j, \Sigma, \Omega_j)$ $(j = 1, \cdots, k)$ 且它们相互独立，那么

$$A = \sum_{j=1}^{k} A_j \sim W_p(n, \Sigma, \Omega)$$

其中 $n = \sum_{j=1}^{k} n_j$，$\Omega = \sum_{j=1}^{k} \Omega_j$。

证明 从定理 2.12 知道 A_j 的特征函数是

$$\varphi_{A_j}(T) = |I_p - 2i\Sigma T|^{-\frac{n_j}{2}} \operatorname{etr}[iT(I_p - 2i\Sigma T)^{-1}\Omega_j]$$

而 A_j 相互独立,所以 $A = \sum_{j=1}^{k} A_j$ 的特征函数为

$$\varphi_A(T) = \prod_{j=1}^{k} \varphi_{A_j}(T) = |I_p - 2i\Sigma T|^{-\frac{n}{2}} \operatorname{etr}[iT(I_p - 2i\Sigma T)^{-1}\Omega],$$

即 $A \sim W_p(n, \Sigma, \Omega)$. ▌

为了进一步理解 Wishart 分布,我们设 $A \sim W_p(n, I_p)$, $n \geqslant p$, 于是 A 概率为 1 正定. 按第一章定理 3.12,知它有唯一的三角分解,即

$$A = TT'$$

其中 T 是下三角矩阵且 $t_{ii} > 0(i = 1, \cdots, p)$, 那么有

定理 2.14 设 $A \sim W_p(n, I_p)(n \geqslant p)$, $A = T \cdot T'$, 其中 $T = (t_{ij})$ 是下三角矩阵且 $t_{ii} > 0(i = 1, \cdots, p)$, 则所有 $t_{ij}(1 \leqslant i \leqslant j \leqslant p)$ 都相互独立,且

$$t_{ij} \sim N(0, 1), \ (i \neq j, 1 \leqslant j < i \leqslant p)$$

$$t_{ii}^2 \sim \chi^2(n - i + 1), \ (i = 1, \cdots, p)$$

证明 见本章附录.

定理 2.15 设 $A \sim W_p(n, I_p)$ $(n \geqslant p)$, 那么

$$|A| = \prod_{i=1}^{p} u_i$$

其中 $u_i \sim \chi^2(n - i + 1)$ 且相互独立.

证明 从定理 2.14 立即可知 $|A| = \prod_{i=1}^{p} t_{ii}^2$ 且 $t_{ii}^2 = u_i \sim \chi^2(n - i + 1)$ 且相互独立. 定理立即得证. ▌

附注: 本定理也可如下证明(不利用定理 2.14):

设

$$A=\begin{vmatrix} A_{11} & \boldsymbol{a}_1 \\ \boldsymbol{a}_1' & a_{pp} \end{vmatrix}$$

其中 A_{11} 是 $(p-1)\times(p-1)$ 矩阵. 于是 $|A|=|A_{11}||a_{pp}-\boldsymbol{a}_1'A_{11}^{-1}\boldsymbol{a}_1|$. 由定理 2.6 知 $a_{pp}-\boldsymbol{a}_1'A_{11}^{-1}\boldsymbol{a}_1\sim W_1(n-p+1,1)$, 即 $\chi^2(n-p+1)$ 且与 $A_{11}\sim W_{p-1}(n,I)$ 独立. 同理将 A_{11} 进行划分. 可知 $|A|=\prod_{i=1}^{p}u_i$, 其中 $u_i\sim\chi^2(n-i+1)$ 且相互独立.

推论 2.7 设 $W\sim W_p(n,\Sigma), n\geqslant p, \Sigma>0$. 则

$$|W|=|\Sigma|\cdot\prod_{i=1}^{p}u_i$$

其中 u_i 相互独立且 $u_i\sim\chi^2(n-i+1)$.

证明 由 $W\sim W_p(n,\Sigma),\Sigma>0$, $n\geqslant p$, 知 $\Sigma^{-\frac{1}{2}}$ 存在, 于是令 $A=\Sigma^{-\frac{1}{2}}W\Sigma^{\frac{1}{2}}$, 按定理 2.7 有 $A\sim W_p(n,I_n)$, 由此可知, $|W|=|\Sigma||A|$. 利用定理 2.15 推得结论. ▮

§3.3 Hotelling T^2 统计量和 Wilks Λ 统计量的分布

在本章我们将介绍另两个重要的多元分布, 即 Hotelling T^2 统计量和 Wilks Λ 统计量的分布, 前者可看成是 t 分布的多元推广, 而后者与 F 分布有着密切的关系.

定义 3.1 设 $A\sim W_p(n,\Sigma)(n\geqslant p,\Sigma>0)$ 和 $\boldsymbol{y}\sim N_p(\boldsymbol{\mu},\Sigma)$ 相互独立, 则称

$$T^2=n\boldsymbol{y}'A^{-1}\boldsymbol{y}$$

为具有自由度 p 和 n 的 Hotelling T^2 变量

T^2 统计量是 Hotelling 最早在讨论多元正态分布的期望向量的检验时提出来的(我们将在第四章中进行讨论)故称为 Hotelling T^2 统计量. 下面我们首先导出它的分布.

定理 3.1

$$\frac{n-p+1}{p}\cdot\frac{T^2}{n}$$

$$=\frac{n-p+1}{p}y'A^{-1}y\sim F(p,n-p+1,\lambda)$$

其中 $\lambda=\mu'\Sigma^{-1}\mu$.

证明　由 $A\sim W_p(n,\Sigma)$, $y\sim N_p(\mu,\Sigma),\Sigma>0,n\geqslant p$ 及引理 2.1 知道 $A^{-\frac{1}{2}}$ 概率为 1 存在,作下列变换

$$\begin{cases}z=\Sigma^{-\frac{1}{2}}y\\W=\Sigma^{-\frac{1}{2}}A\Sigma^{-\frac{1}{2}}\end{cases}$$

其中 $\Sigma^{-\frac{1}{2}}$ 是 Σ^{-1} 的对称平方根,于是 $z\sim N_p(\Sigma^{-\frac{1}{2}}\mu,I_p)$, $W\sim W_p(n,I_p)$ 且

$$y'A^{-1}y=z'W^{-1}z$$

设 P 是一个随机正交矩阵,它的最后一行为 $(z'z)^{-\frac{1}{2}}z'$ 令

$$u=Pz=(0,\cdots0,(z'z)^{\frac{1}{2}}),\ W^*=PWP'.$$

从而当 P 给定后 W^* 的条件分布仍为 $W_p(n,I_p)$, 它与 P 无关, 即 W^* 的无条件分布仍为 $W_p(n,I_p)$ 且与 z 独立, 因此

$$T^2=ny'A^{-1}y=nz'W^{-1}z=nu'(W^*)^{-1}u,$$

记

$$W^*=\begin{pmatrix}W_{11}^* & w_1^*\\ w_1^{*\prime} & w_{pp}\end{pmatrix}\text{ 和 }W^{*-1}=(\omega^{ij})$$

由第一章定理 2.1 和本章定理 2.6 知

$(\omega^{pp})^{-1}=w_{pp}-w_1^{*\prime}W_{11}^{*-1}w_1^*\sim\chi^2(n-p+1)$ 仍和 z 相互独立. 于是

$$T^2=n(0,\cdots,0,\sqrt{z'z})\begin{pmatrix}\omega^{11} & \cdots & \omega^{1p}\\ \omega^{p1} & \cdots & \omega^{pp}\end{pmatrix}\begin{pmatrix}0\\ \vdots\\ 0\\ \sqrt{z'z}\end{pmatrix}$$

$$=n\cdot\frac{z'z}{(\omega^{pp})^{-1}}$$

由于 $z'z \sim \chi^2(p, \mu'\Sigma\mu)$ 与 $(\omega^{pp})^{-1}$ 相互独立，按 F 分布定义知

$$\frac{n-p+1}{p}\frac{T^2}{n} \sim F(p, n-p+1, \lambda)$$

$$\lambda = \mu'\Sigma^{-1}\mu. \quad \blacksquare$$

定理 3.2 设 $x_1, \cdots, x_n$ 是从正态总体 $N_p(\mu, \Sigma)$ 中独立抽取的容量为 n 的样本 $(n > p, \Sigma > 0)$，那么

$$T^2 = n(n-1)\bar{x}'A\bar{x}$$

服从自由度为 p 和 $n-1$ 的 Hotelling T^2 分布，从而

$$\frac{n-p}{p} \cdot \frac{T^2}{n-1} \sim F(p, n-p, \lambda)$$

其中 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$，$A = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'$，$\lambda = n\mu'\Sigma^{-1}\mu$.

证明 根据定理 2.3 知道 $\bar{x} \sim N_p\left(\mu, \frac{1}{n}\Sigma\right)$，$A \sim W_p(n-1, \Sigma)$且它们相互独立，从而 $\sqrt{n}\,\bar{x} \sim N_p(\sqrt{n}\,\mu, \Sigma)$，按 T^2 统计量的定义以及定理 3.1 立即得到所要的结论。 $\blacksquare$

在定理 3.2 中令 $p = 1$，即 T^2 统计量正是通常一元统计中 t 统计量的平方。

定义 3.2 设 $A \sim W_p(n, \Sigma)(\Sigma > 0, n \geqslant p)$，$B \sim W_p(l, \Sigma)$，且 A 与 B 相互独立，则我们称随机变量

$$\Lambda = \frac{|A|}{|A+B|}$$

为 Wilks Λ 变量，它服从的分布记为 $\Lambda(p, n, l)$.

注 在定义 3.2 中要求 $\Sigma > 0$ 和 $n \geqslant p$ 以保证 A 概率为 1 正定，而对 l 没有 $l \geqslant p$ 的要求，当 $p = 1$ 时它正是一元统计中的 β 分布，它最早是 Wilks 提出的。当 $p \geqslant 3$ 时 $\Lambda(p, n, l)$ 的精确分布的密度表达式是很复杂的。

引理 3.1 如果 $\{\mu_i, i = 1, 2\cdots\}$ 是一实数序列，满足

$$\sum_{k=1}^{\infty} \mu_{2k}^{-\frac{1}{2k}}$$

是发散的，那么只有一个分布具有矩序列 $\{\mu_i\}$. 这引理又称为 Carleman 引理，我们将不加证明地予以引用.

设 $\Lambda \sim \Lambda(p,n,l)$ 则 $0 \leqslant \Lambda \leqslant 1$, $0 \leqslant E\Lambda^h = \mu_h \leqslant 1$, 显然满足这个引理，从而它的各阶矩唯一定出了 Λ 的分布. 通过计算 Λ 的 h 阶矩我们可以证明以下定理.

定理 3.3 $\Lambda(p,n,l)$ 的分布正是由 p 个相互独立且依次服从参数为 $\frac{1}{2}(n-i+1)$, $\frac{1}{2}l$ 的 β 分布的随机变数的乘积所服从的分布 $(i=1,\cdots,p)$.

证明见本章附录.

定理 3.4 $\Lambda(p, n,l)$ 和 $\Lambda(l,n+l-p, p)$ 服从相同的分布.

证明见本章附录.

利用定理 3.3 和 3.4 我们可以求出当 $p=1,2$ 或 $l=1,2$ 时 $\Lambda(p,n,l)$ 的确切分布.

1) 当 $p=1$ 时, $\Lambda(1,n,l) \sim \beta\left(\frac{n}{2}, \frac{l}{2}\right)$, 于是由 (1.14) 式知道

$$\frac{n}{l}\frac{1-\Lambda(1,n,l)}{\Lambda(1,n,l)} \sim F(l,n) \tag{3.2}$$

从定理 3.4 知当 $l=1$ 时

$$\begin{aligned}\frac{n+1-p}{p}\frac{1-\Lambda(p,n,1)}{\Lambda(p,n,1)} &= \frac{n-p+1}{p}\frac{1-\Lambda(1,n-p+1,p)}{\Lambda(1,n-p+1,p)} \\ &\sim F(p,n-p+1)\end{aligned} \tag{3.3}$$

2) 当 $p=2$ 时, $\Lambda(2,n,l)=x_1 \cdot x_2$, 其中 $x_1 \sim \beta\left(\frac{n}{2},\frac{l}{2}\right)$, $x_2 \sim \beta\left(\frac{n-1}{2}, \frac{l}{2}\right)$ 且相互独立，于是利用(1.11)式

$$EA^h(2,n,l) = Ex_1^h \cdot Ex_2^h$$

$$= \prod_{i=1}^{2} \frac{\Gamma\left(\frac{n-i+1}{2}+h\right)\Gamma\left(\frac{n-i+1}{2}+\frac{l}{2}\right)}{\Gamma\left(\frac{n-i+1}{2}\right)\Gamma\left(\frac{n-i+1}{2}+\frac{l}{2}+h\right)}$$

利用Γ函数的倍值公式 $\Gamma\left(r+\frac{1}{2}\right)\Gamma(r+1)=2^{-2r}\sqrt{\pi}\,\Gamma(2r+1)$ 可得到

$$EA^h(2,n,l)=\frac{\Gamma(n-1+2h)\Gamma(n+l-1)}{\Gamma(n-1)\Gamma(n+l-1+2h)}$$

$$=E[(y^2)^h] \quad (h=1,2,\cdots)$$

其中 $y \sim \beta(n-1,l)$，由引理 3.1 知$\sqrt{\Lambda(2,n,l)}$ 与$\beta(n-1,\ l)$ 有相同分布，于是有

$$\frac{n-1}{l}\frac{1-\sqrt{\Lambda(2,n,l)}}{\sqrt{\Lambda(2,n,l)}} \sim F(2l,2(n-1)) \qquad (3.4)$$

类似地当 $l=2$ 时有

$$\frac{n-p}{l}\frac{1-\sqrt{\Lambda(p,n,2)}}{\sqrt{\Lambda(p,n,2)}} \sim F(2p,2(n-p)) \qquad (3.5)$$

关于 $p=1,2$ 或者 $l=1,2$ 时Λ的分布与F分布的关系由表(3.1)式给出.

正如F分布对于一元方差分析和回归分析十分有用一样，Λ统计量的分布对于多元方差分析和回归分析也是十分重要的. 关于 $\Lambda(p,n,l)$ 对应于显著水平α的临界值 Λ_α 已经由前人造成表格，读者在使用时可参考文献[9]的书末. 有时为了方便，当n充分大的时候我们也可利用它的渐近分布求出临界域，我们不加证明给出下面定理.

定理 3.5 设 $\Lambda \sim \Lambda(p,n,l)$，则当 $n\to\infty$ 时

$$-r\ln\Lambda \sim \chi^2(pl) \qquad (3.6)$$

其中 $r=n-\frac{1}{2}(p-l+1)$。当 n 不太大时则有

$$P(-r\ln\Lambda\leqslant z)=P(\chi^2(pl)\leqslant z)+\frac{s}{r^2}[P(\chi^2(pl+4)$$

$$\leqslant z)-P(\chi^2(pl)\leqslant z)]+o(n^{-4}) \tag{3.7}$$

其中 $r=n-\frac{1}{2}(p-l+1), s=pl(p^2+l-5)/16$。

此定理是 Box(1949) 给出的，读者可根据显著水平 α 和 n 的大小从 χ^2 表中找到相应的临界值 Λ_α。

关于 Wilks Λ 统计量的渐近分布还有一些其他结果，我们不一一给出了，有兴趣的读者可参考文献[6]。

表 3.1 $\Lambda(p,n,l)$ 与 F 的关系 $(n>p)$

p	l	服从 F 的统计量	自由度
任意	1	$\frac{n-p+1}{p}\frac{1-\Lambda}{\Lambda}$	$(p, n-p+1)$
任意	2	$\frac{n-p}{p}\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}$	$(2p, 2(n-p))$
1	任意	$\frac{n}{l}\frac{1-\Lambda}{\Lambda}$	(l, n)
2	任意	$\frac{n-1}{l}\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}$	$(2l, 2(n-1))$

习 题 三

3.1 设 $y_i=(y_{i1},\cdots,y_{in_i})'\sim N_{n_i}(\mu 1_{n_i}, I_{n_i})$ $(i=1,\cdots,k)$，且相互独立，其中 1_{n_i} 是元素全为 1 的 n_i 维列向量，记

$$n=\sum_{i=1}^{k}n_i$$

$$W=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-y_{\cdot\cdot})^2$$

$$E = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - y_{i\cdot})^2$$

$$B = \sum_{i=1}^{k} n_i (y_{i\cdot} - y_{\cdot\cdot})^2$$

其中 $y_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$，$y_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}$

证明 1) $W = E + B$ 且 $W \sim \chi^2(n-1)$。

2) $E \sim \chi^2(n-k)$，$B \sim \chi^2(k-1)$ 且它们相互独立。

提示：利用习题一中第 8 题的结论。

3.2 设 $z = (x', y')' \sim N_{2p}(0, \Sigma)$，其中 x, y 是 p 维随机向量，而

$$\Sigma = \begin{pmatrix} I_p & \rho I_p \\ \rho I_p & I_p \end{pmatrix}$$

证明 1) $|\rho| \leqslant 1$，当且仅当 $|\rho| < 1$ 时 z 是非奇异分布

2) 令 $Q = I - (y'y)^{-1} yy'$，则

$w = x'Qx \sim (1 - \rho^2) \cdot \chi^2(p-1) \quad (|\rho| < 1)$

提示：首先考虑 y 给定时 w 的条件分布，然后说明它也是无条件分布。

3.3 设 $x_1, \cdots, x_n$ 相互独立且服从 $N(0, \sigma^2)$，证明

$$u = \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2, \quad t = \sum_{i=1}^{n} x_i$$

相互独立，且求出 u 的分布。

提示：设 $x = (x_1, \cdots, x_n)'$，说明 u 可以表示为 x 的二次型 $u = x'Ax$，且 A 是幂等对称阵。

3.4 已知 $A \sim W_p(n, \Sigma)$，记 $A = (a_{ij})$，$\Sigma = (\sigma_{ij})$。证明

1) $EA = n\Sigma$

2) $Ea_{ij}a_{kl} = n^2 \sigma_{ij}\sigma_{kl} + n(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$

3) $E(a_{ij} - n\sigma_{ij})(a_{kl} - n\sigma_{kl}) = n(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$

4) $E(a_{ij} - n\sigma_{ij})^2 = n(\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj})$

提示：利用习题二第 4 题的结论。

3.5 设

$$A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}\sim W_p(n,\Sigma),$$

$$\Sigma=\begin{pmatrix}\Sigma_{11} & 0\\ 0 & \Sigma_{22}\end{pmatrix}>0\ (n\geqslant p)$$

其中 A_{11} 是 $r\times r$ 阶矩阵，A_{22}是 $s\times s$ 阶矩阵且 $r+s=p$，$(r\leqslant s)$ Σ 是相应分块。证明:

$P=A_{12}A_{22}^{-1}A_{21}\sim W_r(s,\Sigma_{11})$ 且与 $(A_{11\cdot 2},A_{22})$ 相互独立。

提示: 利用第三章定理 2.6.

*3.6　多元 t 分布具有密度

$$f(x)=c\cdot|\Sigma|^{-\frac{1}{2}}[1+r^{-1}(x-\mu)'\Sigma^{-1}(x-\mu)]^{-\frac{1}{2}(p+r)}\quad x\in R^p$$

利用推论 2.1 求出常数 c。

*3.7　设 A 是一个 $p\times p$ 阶对称随机矩阵，如果它的概率密度是由下式给出的

$$f(A)=\begin{cases}c|V|^{\frac{N-p-1}{2}}|A|^{-\frac{N}{2}}\operatorname{etr}\left(-\frac{1}{2}VA^{-1}\right) & A\text{ 正定时}\\ 0 & \text{其他}\end{cases}$$

其中 $c=\left[2^{p(N-p-1)}\pi^{\frac{p(p-1)}{4}}\prod_{i=1}^{p}\Gamma\left(\frac{N-p-i}{2}\right)\right]^{-1}$ $(N>2p)$，V 是 $p\times p$ 阶正定矩阵，则称 A 服从自由度为 N，参数矩阵为 V 的逆 Wishart 分布，并记为 $A\sim W_p^{-1}(N,V)$。试证: 当 $S\sim W_p(n,\Sigma)$ 时，则 $S^{-1}\sim W_p^{-1}(N+p+1,\Sigma^{-1})$。

提示: 利用第一章定理 4.5 中 $J(V\to V^{-1})$ 的结论。

*3.9　设 $W_1\sim W_p(n_1,I)$，$W_2\sim W_p(n_2,I)(n_1\geqslant p,n_2\geqslant p)$。

证明: $Z=W_1^{-\frac{1}{2}}W_2W_1^{-\frac{1}{2}}$的分布密度为

$$c\cdot|Z|^{\frac{1}{2}(n_2-p-1)}|I+Z|^{\frac{1}{2}(n_1+n_2)}$$

并求出常数 c，并说明为什么当 $W_i\sim W_p(n_i,\Sigma)(\Sigma\neq I)$ 时不能得出以上结论。

提示: 利用 $J(W_1,W_2\to Z,W_2)=|W_1|^{\frac{1}{2}(p+1)}$。

*3.10　已知 X 是 $p\times n$ 阶随机矩阵，$(n\geqslant p)$，且它的密度为

$$f(X)=c\cdot|XX'|^r\operatorname{etr}\left(-\frac{1}{2}XX'\right)\quad(r\text{ 常数})$$

求 $S = XX'$ 的密度函数，并求出常数 c。

提示：利用推论 1.2。

*3.11 设 $X \sim N_{p,n}(M, \Sigma, V)$。 $Y_i = XA_iX'$ $(i = 1, \cdots, k)$，$Y = XAX' = \sum_{i=1}^{k} Y_i$。若

1) $Y, Y_1, \cdots, Y_{k-1}$ 都服从非中心 Wishart 分布，

2) 对所有 X，$Y_k \geq 0$。

证明 $Y_1, \cdots, Y_k$ 都相互独立且服从非中心 Wishart 分布。

提示：利用第三章引理 1.1。

附　录

(一) 引理 1.2 的证明

1) 由条件 1)，2)推出条件 3)，4)。

由 A_i 幂等和 $A_i \cdot A_j = 0$ $(i \neq j,\ i, j = 1, \cdots, k)$ 知道

$$A^2 = \left(\sum_{i=1}^{k} A_i\right)^2 = \sum_{i=1}^{k} A_i^2 + \sum_{i \neq j} A_i \cdot A_j = \sum_{i=1}^{k} A_i = A$$

因此 A 幂等。另外

$$p = \mathrm{rk}(A) = \mathrm{tr}A = \mathrm{tr}\left(\sum_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \mathrm{tr}(A_i)$$

$$= \sum_{i=1}^{k} \mathrm{rk}(A_i) = \sum_{i=1}^{k} r_i.$$

2) 由条件 1)，3)推出条件 2)。

由 A_i 及 A 幂等，知道 A_k 是非负定阵，因此存在矩阵 B 使得 $A_k = BB'$，由引理 1.1 立即可得 2)：$A_i \cdot A_j = 0$ $(i \neq j; i, j = 1, \cdots, k)$。

3) 由条件 2)，3)推出条件 1)。

由 $A_i \cdot A_j = 0$ $(i \neq j, i, j = 1, \cdots, k)$ 及 A 幂等，知

$$A_i^2 = A_i A = A_i A^2 = A_i^3.$$

又因为

$$\begin{pmatrix} I & I \\ 0 & I \end{pmatrix}\begin{pmatrix} 0 & A_i - A_i^2 \\ A_i^2 & 0 \end{pmatrix}\begin{pmatrix} I & 0 \\ I & I \end{pmatrix} = \begin{pmatrix} A_i & A_i - A_i^2 \\ A_i^2 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} I & 0 \\ A_i & I \end{pmatrix}\begin{pmatrix} A_i & 0 \\ 0 & -A_i^2(I - A_i) \end{pmatrix}\begin{pmatrix} I & I - A_i \\ 0 & I \end{pmatrix}$$

所以 $$\mathrm{rk}(A_i - A_i^2) + \mathrm{rk}(A_i^2) = \mathrm{rk}(A_i) + \mathrm{rk}(A_i^2 - A_i^2)$$

而 $\mathrm{rk}(A_i^2) = \mathrm{rk}(A_i \cdot A_i') = \mathrm{rk}(A_i)$，代入上式，得到

$$\mathrm{rk}(A_i - A_i^2) = 0,$$

即 $A_i = A_i^2\ (i = 1, \cdots, k)$，1）式成立。

4）由条件 3)，4)推出条件 1)。

由于 A 幂等，知道存在正交阵 P 使得

$$P'AP = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix}$$

记

$$P'A_iP = \begin{pmatrix} D_i & F_i' \\ F_i & G_i \end{pmatrix},$$

则

$$\begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} = \sum_{i=1}^{k} \begin{pmatrix} D_i & F_i' \\ F_i & G_i \end{pmatrix} \tag{1}$$

由 $I_p = \sum_{i=1}^{k} D_i$ 以及 $p = \sum_{i=1}^{k} r_i$ 得到

$$p = \mathrm{rk}(I_p) \leqslant \sum_{i=1}^{k} \mathrm{rk}(D_i) \leqslant \sum_{i=1}^{k} \mathrm{rk}(A_i) = \sum_{i=1}^{k} r_i = p$$

所以 $r_i = \mathrm{rk}(D_i) = \mathrm{rk}(A_i)$。

下面证明 D_i 是幂等阵。由 D_1 对称，存在正交阵 Q 使得

$$Q'D_1Q = \mathrm{diag}(\lambda_1, \cdots \lambda_{r_1}, 0, \cdots, 0)$$

于是

$$I_p = \mathrm{diag}(\lambda_1, \cdots \lambda_{r_1}, 0, \cdots, 0) + \sum_{i=2}^{k} Q'D_iQ$$

因此

$$\mathrm{rk}\begin{pmatrix} 1-\lambda_1 & & & & \\ & \ddots & & & \\ & & 1-\lambda_{r_1} & & \\ & & & 1 & \\ & & & & \ddots \\ & & & & & 1 \end{pmatrix} \leqslant \sum_{i=2}^{k} \mathrm{rk}(Q'D_iQ)$$

$$= \sum_{i=2}^{k} \mathrm{rk}(D_i) = p - r_1$$

必有 $\lambda_1 = \cdots = \lambda_{r_1} = 1$，于是 D_1 是幂等对称阵，类似可以证明 $D_i(i = 2,\cdots,k)$ 也是幂等矩阵。

最后证明 $G_i = 0$，$F_i = 0$。由于

$$\mathrm{rk}\begin{pmatrix} D_i & F_i' \\ F_i & G_i \end{pmatrix} = \mathrm{rk}(D_i)$$

存在矩阵 T 使得

$$(F_i,\ G_i) = T(D_i,\ F_i')$$

所以 $G_i = TF_i' = TD_iT' \geqslant 0$。但从(1)式知 $\sum_{i=1}^{k} G_i = 0$，推得 $G_i = 0(i = 1,\cdots,k)$。又 $0 = G_i = TD_iT' = (TD_i)(TD_i)'$
即 $F_i = TD_i = 0\ (i = 1,\cdots k)$。最后推得

$$A_i = P\begin{pmatrix} D_i & 0 \\ 0 & 0 \end{pmatrix}P'$$

是幂等阵，所以条件 1) 成立。 ▮

（二）定理 2.14 的证明。

由于 $A \sim W_p(n, I_p)(n \geqslant p)$，$A$ 有密度

$$f(A) = c(p,n)|A|^{\frac{1}{2}(n-p-1)}\mathrm{etr}\left(-\frac{1}{2}A\right) \tag{2}$$

其中
$$c(p,n) = \left[2^{\frac{1}{2}pn}\Gamma_p\left(\frac{n}{2}\right)\right]^{-1},$$

$$\Gamma_p\left(\frac{n}{2}\right) = \pi^{\frac{1}{4}p(p-1)}\prod_{i=1}^{p}\Gamma\left(\frac{n-i+1}{2}\right). \tag{3}$$

作变换 $A = TT'$，其中 T 是对角元素为正的下三角矩阵，按第一章定理 4.4 知

$$J(A \to T) = 2^p \cdot \prod_{i=1}^{p} t_{ii}^{p-i+1}$$

代入(2)式，得到 T 的密度为

$$f(T) = c(p,n)\cdot 2^p \cdot \prod_{i=1}^{p} t_{ii}^{p-i+1}\left(\prod_{i=1}^{p} t_{ii}^2\right)^{\frac{1}{2}(n-p-1)}$$

$$\cdot \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{p} t_{ii}^2 + \sum_{i<j}\sum t_{ij}^2\right)\right]$$

从因子分解定理知道所有 $t_{ij}(1\leqslant i<j\leqslant p)$ 相互独立且 $f(T)$ 可写成

$$f(T)=\prod_{1\leqslant i<j\leqslant p}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t_{ij}^2\right)$$

$$\cdot\prod_{i=1}^{p}\frac{2t_{ii}}{\Gamma\left(\frac{n-i+1}{2}\right)\cdot 2^{\frac{n-i+1}{2}}}(t_{ii}^2)^{\frac{n-i+1}{2}}$$

$$\cdot\exp\left(-\frac{1}{2}t_{ii}^2\right)\tag{4}$$

即 $t_{ii}^2\sim\chi^2(n-i+1)(i=1,\cdots,p)$ 以及 $t_{ij}\sim N(0,1)(1\leqslant i<j\leqslant p)$。 ▮

（三）定理 3.3 的证明

我们只须证明 $\Lambda\sim\Lambda(p,n,l)$ 的 h 阶矩恰等于 p 个独立且服从

$$\beta\left(\frac{n-i+1}{2},\ \frac{l}{2}\right)\quad(i=1,\cdots,p)$$

的随机变数乘积的 h 阶矩即可。

容易证明 Λ 的分布与 Σ 无关，不妨设 $A\sim W_p(n,I_p)$ 和 $B\sim W_p(l,I_p)$ 且相互独立，$n\geqslant p$。于是 $B=YY'$，其中

$Y\sim N_{p,l}(0,I_p,I_l)$ 且与 A 相互独立。

于是 (A,Y) 的联合密度为

$$f(A,Y)=c(p,n)\cdot|A|^{\frac{n-p-1}{2}}(2\pi)^{-\frac{pl}{2}}\operatorname{etr}\left[-\frac{1}{2}(A+YY')\right]\tag{5}$$

因此

$$E\Lambda^h=\int\cdots\int|A|^h|A+YY'|^{-h}f(A,Y)dAdY$$

$$=\frac{c(p,n)}{c(p,n+2h)}\int\cdots\int|A+YY'|^{-h}c(p,n+2h)$$

$$\cdot|A|^{\frac{n+2h-p-1}{2}}(2\pi)^{-\frac{pl}{2}}$$

$$\cdot\operatorname{etr}\left[-\frac{1}{2}(A+YY')\right]dAdY$$

注意到积分号下可看成 $A\sim W_p(n+2h,I_p)$ $YY'\sim W_p(l,I_p)$ 且相互独立，则 $G=A+YY'\sim W_p(n+2h+l,I_p)$，此时 $E\Lambda^h$ 可借助于 $E|G|^{-h}$ 得到，而

$$E|G|^{-h}=\int_{G>0}|G|^{-h}\cdot c(p,n+2h+l)|G|^{\frac{n+2h+l-p-1}{2}}$$

$$\cdot\operatorname{etr}\left(-\frac{1}{2}G\right)dG$$

$$= \frac{c(p, n+2h+l)}{c(p, n+l)} \int_{G>0} c(p, n+l) |G|^{\frac{n+l-p-1}{2}}$$

$$\cdot \operatorname{etr}\left(-\frac{1}{2} G\right) dG$$

$$= \frac{c(p, n+2h+l)}{c(p, n+l)}.$$

于是

$$E\Lambda^h = \frac{c(p, n)}{c(p, n+2h)} E|G|^{-h} = \frac{c(p, n) c(p, n+2h+l)}{c(p, n+2h) c(p, n+l)}$$

$$= \frac{\Gamma_p \frac{n+l}{2} \cdot \Gamma_p \left(\frac{n}{2} + h\right)}{\Gamma_p \left(\frac{n}{2}\right) \cdot \Gamma_p \left(\frac{n+l}{2} + h\right)}$$

$$= \prod_{i=1}^{p} \frac{\Gamma\left(\frac{n+1-i}{2} + \frac{l}{2}\right) \Gamma\left(\frac{n+1-i}{2} + h\right)}{\Gamma\left(\frac{n+1-i}{2}\right) \Gamma\left(\frac{n+1-i}{2} + h + \frac{l}{2}\right)}. \tag{6}$$

这是将(3)式代入后计算出来的.

按本章(1.11)式知,若 $x_i \sim \beta\left(\frac{n+1-i}{2}, \frac{l}{2}\right)$时,它的 h 阶矩正是(6)式中的一个因子,而 x_i 相互独立,则它们乘积的 h 阶矩正是各个变数 h 阶矩的乘积,从而定理得证. ∎

(四) 定理 3.4 的证明

设 $\Lambda_1 \sim \Lambda(p, n, l)$, $\Lambda_2 \sim \Lambda(l, n+l-p, p)$, 要让它们具有相同的分布,只须证明 $E\Lambda_1^h = E\Lambda_2^h$ $(h = 1, 2, \ldots)$即可.

当 $l \leqslant p$ 时 ($l \geqslant p$ 可类似证明). 由于

$$\Gamma_p\left(\frac{n}{2}\right) = \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left(\frac{n+1-i}{2}\right)$$

$$\Gamma_l\left(\frac{n}{2}\right) = \pi^{\frac{l(l-1)}{4}} \prod_{i=1}^{l} \Gamma\left(\frac{n+1-i}{2}\right)$$

$$\Gamma_{p-l}\left(\frac{n-l}{2}\right) = \pi^{\frac{(p-l)(p-l-1)}{4}} \prod_{i=1}^{p-l} \Gamma\left(\frac{n-l-i+1}{2}\right)$$

从而

$$\Gamma_p\left(\frac{n}{2}\right)=\pi^{\frac{1}{2}l(p-l)}\Gamma_l\left(\frac{n+l-p}{2}\right)\Gamma_{p-l}\left(\frac{n}{2}\right)$$

$$=\pi^{\frac{1}{2}l(p-l)}\Gamma_l\left(\frac{n}{2}\right)\Gamma_{p-l}\left(\frac{n-l}{2}\right).$$

在上式中以 $n+l$ 代替 n 后可得到

$$\frac{\Gamma_p\left(\frac{n+l}{2}\right)}{\Gamma_p\left(\frac{n}{2}\right)}=\frac{\Gamma_l\left(\frac{n+l}{2}\right)}{\Gamma_l\left(\frac{n+l-p}{2}\right)} \tag{7}$$

由(6)式知

$$EA_1^h=\frac{\Gamma_p\left(\frac{n+l}{2}\right)\Gamma_p\left(\frac{n}{2}+h\right)}{\Gamma_p\left(\frac{n}{2}\right)\Gamma_p\left(\frac{n+l}{2}+h\right)}$$

$$EA_2^h=\frac{\Gamma_l\left(\frac{n+l}{2}\right)\Gamma_l\left(\frac{n+l-p}{2}+h\right)}{\Gamma_l\left(\frac{n+l-p}{2}\right)\Gamma_l\left(\frac{n+l}{2}+h\right)} \tag{8}$$

将(7)式代入(8)式，并用 $n+2h$ 代替 n，即得 $EA_1^h=EA_2^h$。 ▌

第四章　多元正态分布参数的估计

若 p 维随机向量 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$，那么它的分布密度完全由它的期望向量 $\boldsymbol{\mu}$ 和协方差矩阵 Σ 所决定．因而当已知某一随机向量服从正态分布时，如何通过抽取的随机样本来估计它的数学期望向量 $\boldsymbol{\mu}$ 和协方差矩阵 Σ 是一个十分重要的问题．与其相联系，对于多元正态分布的其他参数，如相关系数，偏相关系数，广义方差等也有相应的估计问题．此外由于这些估计量都是样本的统计量，当然希望了解它们所服从的分布，从而为进一步探讨这些统计量与被估计的参数的关系创造条件．本章要讨论的正是多元正态分布的参数估计问题．

在一元统计分析中，大家都已熟知求参数的极大似然估计的方法．本章主要的内容就是利用极大似然法求出多元正态分布各种参数的极大似然估计．由于极大似然估计往往比较简单而且具有许多优良性质，在实际问题中得到许多应用．

在本章开始，我们先将有关记号作一叙述．

设有某一个 p 维正态总体 $\sim N_p(\boldsymbol{\mu},\Sigma)$，$\Sigma > 0$，对决定该分布的参数 $\boldsymbol{\mu}$ 和 Σ 我们要作出估计．为此我们从该总体中独立地抽取了容量为 n 的随机样本(通常 $n > p$)，即简单样本，记为 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$．我们将在 $\boldsymbol{x}_i$ 的观察值为基础上进行整理，分析，然后得到 $\boldsymbol{\mu}$，Σ 的估计．以后我们记

$$X = (\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n) \qquad 为样本矩阵$$

$$\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i = \frac{1}{n} X \cdot \mathbf{1}_n \qquad 为样本均值向量 \tag{1.1}$$

$$S = \frac{1}{n-1} A \qquad 为样本协方差矩阵$$

其中

$$A = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})' = (X - \bar{x}1_n')(X - \bar{x}1_n')'$$

$$= XX' - n\bar{x}\bar{x}' \qquad = X\left(I_n - \frac{1}{n}1_n1_n'\right)X' \quad (1.2)$$

相应的我们称$|\Sigma|$为正态总体的"广义方差"，而$|S|$则称为样本的"广义方差"。

如同总体协方差矩阵Σ的分块一样，我们也常将S（或A）进行相应的分块，若记

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

其中S_{11}是$r \times r$阶，S_{22}是$s \times s$阶，$r + s = p$。则

S_{11}, S_{22} 称样本的子协方差矩阵

$S_{1.2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$，$S_{2.1} = S_{22} - S_{21}S_{11}^{-1}S_{12}$ 称样本的偏（条件）协方差矩阵，

引用第二章第四节的记号，则有

$$X \sim N_{p,n}(\mu 1_n', \Sigma, I_n)$$

而相应于$n > p$时样本的联合密度为

$$f(X) = (2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}\text{etr}\left[-\frac{1}{2}\Sigma^{-1}(X - \mu 1_n')(X - \mu 1_n')'\right]$$

$$= (2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}\text{etr}\left[-\frac{1}{2}\Sigma^{-1}(A + n(\bar{x} - \mu)(\bar{x} - \mu)'\right]$$

(1.3)

其中第二等式是在$X - \mu 1_n'$中分别加减$\bar{x}1_n'$计算得到的。

§4.1 期望向量μ和协方差矩阵Σ的估计

在本节中我们将讨论p维正态总体$N_p(\mu, \Sigma)$的参数μ和Σ的极大似然估计，一致最小方差无偏估计以及这些估计量的渐近

正态性等性质.

(一) μ 及 Σ 的极大似然估计

首先我们用极大似然法求出 $N_p(\mu,\Sigma)$ 中 μ 和 Σ 的极大似然估计,然后讨论这些估计量的性质及其分布.

设 $x_1,\cdots,x_n$ 相互独立且服从 $N_p(\mu,\Sigma),\Sigma>0,n>p$, 即 $X=(x_1,\cdots,x_n)\sim N_{p,n}(\mu 1_n',\Sigma,I_n)$. 此时 X 的联合密度为(1.3)式. 它既是 X 的函数, 又是 μ 和 Σ 的函数, 在寻求 μ 和 Σ 的极大似然估计时, 在(1.3)中将 X 的值看成是固定的仅将(1.3)视为(μ, Σ)的函数,此时(1.3)称为似然函数 $L(\mu,\Sigma)$. 注意到(1.3)式的两重意义的差别. 下面求 $L(\mu,\Sigma)$ 的极值.

引理 1.1 设 S 是 $p\times p$ 阶正定矩阵,那么对于所有 $p\times p$ 阶正定矩阵 Σ,有

$$f(\Sigma)=|\Sigma|^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\operatorname{tr}\Sigma^{-1}S\right\}\leqslant\left|\frac{S}{n}\right|^{-\frac{n}{2}}\exp\left(-\frac{np}{2}\right)\qquad(1.4)$$

且等号成立的充分必要条件是 $\Sigma=\frac{1}{n}S$.

证明 由于 Σ,S 正定,按第一章定理 3.11, 存在非奇异矩阵 P,使得 $S=PP'$, $\Sigma=PD_\lambda P'$, 其中 $D_\lambda=\operatorname{diag}(\lambda_1,\cdots,\lambda_p)$, λ_i 是 Σ 相对于 S 的特征根,于是当 Σ 变动时 λ_i 也随之变动,由于

$$|\Sigma|=|PD_\lambda P'|=|S|\cdot\prod_{i=1}^{p}\lambda_i$$

$$\operatorname{tr}\Sigma^{-1}S=\operatorname{tr}[(P')^{-1}D_\lambda^{-1}P^{-1}\cdot P\cdot P']=\operatorname{tr}D_\lambda^{-1}=\sum_{i=1}^{p}\lambda_i^{-1}$$

(1.4)式可改写为

$$\begin{aligned}f(\lambda_1,\cdots,\lambda_p)&=|S|^{-\frac{n}{2}}\left(\prod_{i=1}^{p}\lambda_i\right)^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{p}\lambda_i^{-1}\right)\\&=|S|^{-\frac{n}{2}}\prod_{i=1}^{p}\left[\lambda_i^{-\frac{n}{2}}\exp\left(-\frac{1}{2\lambda_i}\right)\right]\qquad(1.5)\end{aligned}$$

由对数函数的单调性,(1.3)或(1.5)极大等价于对每一 i 使得

$$g(\lambda_i) = -\frac{n}{2}\ln\lambda_i - \frac{1}{2\lambda_i}$$

达到最大。由于当 $\lambda_i \gtrless \frac{1}{n}$ 时有

$$g'(\lambda_i) = \frac{1-n\lambda_i}{2\lambda_i^2} \lesseqgtr 0$$

所以 $g(\lambda_i)$ 当且仅当 $\lambda_i = \frac{1}{n}$ 时达到极大。即

$$f(\Sigma) \leqslant \left|\frac{S}{n}\right|^{-\frac{n}{2}} \exp\left(-\frac{np}{2}\right)$$

且等式成立的充分必要条件是 $\lambda_1 = \cdots = \lambda_p = \frac{1}{n}$，即

$$\Sigma = PD_\lambda P' = \frac{1}{n}PP' = \frac{1}{n}S. \ \blacksquare$$

定理 1.1 当 $X \sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}_n', \Sigma, I_n), \Sigma > 0, n > p$ 时 $\boldsymbol{\mu}$ 及 Σ 的极大似然估计分别为 $\bar{x}$ 和 $\frac{1}{n}A$（$\bar{x}$ 和 A 由 (1.1)，(1.2) 式所定义）。

证明　由(1.3)式知道似然函数为

$$L(\boldsymbol{\mu}, \Sigma) = (2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}\exp\left[-\frac{1}{2}\operatorname{tr}\Sigma^{-1}A - \frac{1}{2}n(\bar{x}-\boldsymbol{\mu})'\Sigma^{-1}(\bar{x}-\boldsymbol{\mu})\right]$$

又由 $\Sigma > 0, n > p$ 及第三章引理 2.1 知道 A 以概率 1 正定，于是由引理 1.1 知道当

$$\hat{\boldsymbol{\mu}} = \bar{x},\ \hat{\Sigma} = \frac{1}{n}A$$

时 $L(\boldsymbol{\mu}, \Sigma)$ 达到它的极大值，即

$$L(\boldsymbol{\mu}, \Sigma) \leqslant L\left(\bar{x}, \frac{1}{n}A\right) = (2\pi)^{-\frac{np}{2}}\left|\frac{1}{n}A\right|^{-\frac{n}{2}}\exp\left(-\frac{np}{2}\right).$$

所以 $\bar{x}$ 和 $\frac{1}{n}A$ 是 $\boldsymbol{\mu}$ 和 Σ 的极大似然估计。 ▮

根据第二章定理 2.3 我们知道

$$\bar{x} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\Sigma\right), \quad A \sim W_p(n-1, \Sigma)$$

且相互独立。

(二) $\boldsymbol{\mu}$ 及 $\boldsymbol{\Sigma}$ 的一致最小方差无偏估计

定理 1.2 $\bar{x}$ 和 $S = \frac{1}{n-1}A$ 是 $\boldsymbol{\mu}$ 和 Σ 的无偏估计

证明 由于 $\bar{x} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\Sigma\right)$，所以 $E\bar{x} = \boldsymbol{\mu}$，$\bar{x}$ 是 $\boldsymbol{\mu}$ 的无偏估计，且此时 $\sqrt{n}(\bar{x}-\boldsymbol{\mu}) \sim N_p(0, \Sigma)$，故

$$\begin{aligned} EA &= E\left[\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{x})(\boldsymbol{x}_i - \bar{x})'\right] \\ &= E\left[\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})' - \sqrt{n}(\bar{x} - \boldsymbol{\mu}) \right. \\ &\quad \left. \cdot \sqrt{n}(\bar{x} - \boldsymbol{\mu})'\right] \\ &= n\Sigma - \Sigma = (n-1)\Sigma. \end{aligned}$$

所以 $ES = \Sigma$，即 $S = \frac{1}{n-1}A$ 是 Σ 的无偏估计。

由前面所述，$(\bar{x}, A)$ 对于 $(\boldsymbol{\mu}, \Sigma)$ 的估计十分重要，不仅如此，我们还将指出 $(\bar{x}, A)$ 是正态分布族的充分完备统计量. 这里我们主要对充分统计量进行讨论. 为此我们简单回顾一下有关概念.

1) 设 $X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$ 是从分布族为 $\{F(\boldsymbol{x}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ 中抽取的样本矩阵，$T(X)$ 是某一统计量，如果给定 $T(X) = t$ 后，X 的条件分布与参数 $\boldsymbol{\theta}$ 无关，则称 $T(X)$ 是分布族 $F(\boldsymbol{x}, \boldsymbol{\theta})$（或 $\boldsymbol{\theta}$）的充分统计量.

按 Neyman 因子分解定理：如果总体分布具有密度 $f(\boldsymbol{x}, \theta)$

且样本的联合密度可分解为

$$f_1(X,\theta) = \prod_{i=1}^{n} f(x_i,\theta) = h(X)\cdot g(T(X),\theta)$$

其中 $T(X)$ 是统计量，$h(X)$ 是样本的非负函数且与 θ 无关，$g(T(X),\theta)$仅通过 $T(X)$ 依赖于样本 X，那末 $T(X)$ 是 θ 的充分统计量。

2）设分布族为 $\{F(x,\theta),\theta\in\Theta\}$，如果对任何一个满足

$$E_\theta g(x)=0 \qquad \text{对一切 } \theta \text{ 成立}$$

的随机变量 $g(x)$，总有

$$P_\theta(g(x)=0)=1 \qquad \text{对一切 } \theta \text{ 成立}$$

则称 $\{F(x,\theta),\theta\in\Theta\}$ 是完备的。如果 $X=(x_1,\cdots,x_n)$ 是从该分布族抽取的样本矩阵，统计量 $T(X)$ 的导出分布族是完备的，那么称 $T(X)$ 是分布族 $\{F(x,\theta)\cdot\theta\in\Theta\}$（或 θ）的完备统计量。

3）设 $X=(x_1,\cdots,x_n)$ 是取自某一总体的样本矩阵，总体参数为 θ，如果存在 θ 的一个无偏估计 $\hat{\theta}(X)$，使得对 θ 的任意一个无偏估计 $\tilde{\theta}(X)$，总有

$$\operatorname{cov}(\hat{\theta}(X)) \leqslant \operatorname{cov}(\tilde{\theta}(X)) \qquad \text{对一切 } \theta \text{ 成立}$$

则称 $\hat{\theta}(X)$ 是 θ 的一致最小方差无偏估计

若 $T(X)$ 是 θ 的充分完备统计量，且 $g(T(X))$ 是 θ 的无偏估计，那么 $g(T(X))$ 是 θ 的一致最小方差无偏估计。

定理 1.3 $(\bar{x},A)$ 是 (μ,Σ) 的充分统计量。

证明　由于样本联合密度为

$$f(X)=(2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}} \cdot \operatorname{etr}\left\{-\frac{1}{2}\Sigma^{-1}[A+n(\bar{x}-\mu)(\bar{x}-\mu)']\right\}$$

按因子分解定理即可知 $(\bar{x},A)$ 是 (μ,Σ) 的充分统计量。 ▌

此外还可证明 $(\bar{x},A)$ 是 (μ,Σ) 的完备统计量（参考 [5]）。

定理 1.4 $\bar{x}$ 和 $S=\frac{1}{n-1}A$ 是 μ 和 Σ 的一致最小方差无

偏估计。

证明　由于 $\bar{x}$ 和 S 是 μ 和 Σ 的无偏估计，而且$(\bar{x}, A)$是充分完备统计量，所以它们是一致最小方差无偏估计。 ▮

由上面讨论知 μ 的极大似然估计 $\bar{x}$ 同时也是一致最小方差无偏估计． 而 Σ 的极大似然估计 $\frac{1}{n}A$ 与一致最小方差无偏估计 $\frac{1}{n-1}A$ 相差一个倍数 $\frac{n}{n-1}$。在实用中常用 S 估计 Σ，因此称 S 为样本的协方差矩阵．

关于 $\bar{x}$ 和 A 的其他优良性质等可参考[6]。

(三)　$\bar{x}$ 和 S 的大样本性质

在(一)，(二)讨论中总假定总体为正态分布，事实上即使总体不是正态分布的，$\bar{x}$ 和 S 当样本容量趋于无限时的仍具有良好的性质。下面定理 1.5 和定理 1.6 分别讨论了它们的按概率收敛性和渐近正态性．

设 $X=(x_1, \cdots x_n)$ 是取自某一总体(不必为正态)的样本矩阵，总体的数学期望向量为 μ，协方差矩阵为 Σ，$\bar{x}$ 和 A 是由 (1.1) 和(1.2)所定义的．那么有

定理 1.5(大数定理)　设 $x_1, \cdots x_n$ 是从具有均值向量 μ，协方差矩阵 Σ 的总体中独立抽取的容量为 n 的简单样本，则当 $n \to \infty$时有

1) $\bar{x}$ 按概率收敛于 μ，即 $\bar{x} \xrightarrow{P} \mu$。　　(1.6)

2) $\frac{1}{n}A$ 按概率收敛于 Σ，即 $\frac{1}{n}A \xrightarrow{P} \Sigma$　　(1.7)

证明　利用一元的大数定理，知道对总体的每一分量有

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij} \xrightarrow{P} \mu_i \quad (i=1, \cdots, p)$$

其中 x_{ij} 是 x_j 的第 i 个分量，$\bar{x}_i$ 是 $\bar{x}$ 的第 i 个分量。从而 1) 成立。

若记 $A=(a_{ij}),\Sigma=(\sigma_{ij}),\boldsymbol{x}_i=(x_{1i},\cdots,x_{pi})',\bar{\boldsymbol{x}}=(\bar{x}_1,\cdots,\bar{x}_p)'$，则

$$\frac{1}{n}a_{ij}=\frac{1}{n}\sum_{k=1}^{n}(x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j)$$

$$=\frac{1}{n}\sum_{k=1}^{n}(x_{ik}-\mu_i+\mu_i-\bar{x}_i)(x_{jk}-\mu_j+\mu_j-\bar{x}_j)$$

$$=\frac{1}{n}\sum_{k=1}^{n}(x_{ik}-\mu_i)(x_{jk}-\mu_j)-(\bar{x}_i-\mu_i)(\bar{x}_j-\mu_j)$$

由于 $E(x_{ik}-\mu_i)(x_{jk}-\mu_j)=\sigma_{ij}$，由一元大数定理知道上式第一项按概率收敛于 σ_{ij}，而第二项按概率收敛于 0，于是有$\frac{1}{n}A\xrightarrow{P}\Sigma$。▮

式(1.6)、(1.7)的实际意义是：当 n 充分大时，$\bar{\boldsymbol{x}}$ 和 $\frac{1}{n}A$ 以接近于 1 的概率收敛于 $\boldsymbol{\mu}$ 及 Σ，而以 S 代替 $\frac{1}{n}A$ 也有相同的结论。

定理 1.6(中心极限定理) 设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 是从具有均值向量 $\boldsymbol{\mu}$ 和协方差矩阵 Σ 的总体中独立抽取的容量为 n 的简单样本，则当 $n\to\infty$时有

$$\sqrt{n}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})$$

渐近地服从 $N_p(\boldsymbol{0},\Sigma)$。

证明 只须证明 $\sqrt{n}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})$ 的特征函数当 $n\to\infty$时收敛于 $N_p(\boldsymbol{0},\Sigma)$ 的特征函数即可。

设 $\boldsymbol{t}$ 是任一 p 维向量，则 $u_n=\boldsymbol{t}'\cdot\sqrt{n}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})$是一维变数且 $Eu_n=0$，$Eu_n^2=\boldsymbol{t}'\Sigma\boldsymbol{t}$。由一元中心极限定理知道 u_n 渐近地服从 $N(0,\boldsymbol{t}'\Sigma\boldsymbol{t})$，而 u_n 的特征函数是

$$\varphi_n(s,\boldsymbol{t})=Ee^{isu_n}=E\exp[is\cdot\boldsymbol{t}'\sqrt{n}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})]$$

而 $N(0,\boldsymbol{t}'\Sigma\boldsymbol{t})$ 相应特征函数是 $\exp\left(-\frac{s^2}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right)$，于是

$$\lim_{n\to\infty} \varphi_n(s,t) = \exp\left(-\frac{s^2}{2}t'\Sigma t\right)$$

令 $s=1$，就得到

$$\lim_{n\to\infty} E[\exp it'\sqrt{n}(\bar{x}-\mu)] = \exp\left\{-\frac{1}{2}t'\Sigma t\right\}$$

即 $\sqrt{n}(\bar{x}-\mu)$ 的特征函数当 $n\to\infty$ 收敛于 $N_p(0,\Sigma)$ 的特征函数。定理得证。 ∎

定理 1.6 告诉我们，即使总体不是正态分布，但它的样本均值向量当容量 n 充分大时总可以看成是正态分布，而且相应的参数 μ 和 Σ 可以用 $\bar{x}$ 和 $\frac{1}{n}A$(或 S)来代替.这样一来，非正态总体只要它的样本容量 n 充分大，都可以用前面讨论的正态总体来研究.例如 $\sqrt{n}(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)$ 当 n 充分大时渐近服从 $\chi^2(p)$ 分布，而当 Σ 未知时，用 S^{-1} 代替 Σ^{-1}，$\sqrt{n}(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu)$ 也可证明它渐近服从 $\chi^2(p)$ 分布。

§4.2 广义方差和相关系数的极大似然估计及它们的分布

本节将讨论正态总体的广义方差 $|\Sigma|$，相关系数 ρ_{ij}，偏相关系数，复相关系数的极大似然估计，并对用作估计的统计量的分布进行较详细讨论，这将为以后几章中的检验等问题的讨论奠定基础。

（一） 广义方差和相关系数的极大似然估计

首先回顾一下第二章第三节中有关概念及记号：

设总体 $x\sim N_p(\mu,\Sigma)$，$\Sigma=(\sigma_{ij})>0$，Σ 可进行如下分块

$$\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\begin{matrix}r\\ s\end{matrix},\quad r+s=p, \text{则}$$

$\rho_{ij}=\sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ 称为分量 x_i 与 x_j 的相关系数。

$\Sigma_{1.2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}=(\sigma_{ij\cdot r+1,\cdots,p})$ 称为偏（条件）协方差

矩阵

$\beta = \Sigma_{12}\Sigma_{22}^{-1}$ 称为回归系数矩阵。

$$\rho_{ij.r+1,\cdots,p} = \frac{\sigma_{ij.r+1,\cdots,p}}{\sqrt{\sigma_{ii.r+1,\cdots,p}\cdot\sigma_{jj.r+1,\cdots,p}}}$$

称为 x_i 和 x_j 的偏相关系数。

若记

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\begin{matrix} 1 \\ p-1 \end{matrix} \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{21}' \\ \sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$\rho_{1.2,\cdots,p} = \left[\frac{\sigma_{21}'\Sigma_{22}^{-1}\sigma_{21}}{\sigma_{11}}\right]^{\frac{1}{2}}$$

称为 x_1 和 $x_2 = (x_2,\cdots,x_p)'$ 之间的复相关系数。

$|\Sigma|$ 称为总体的广义方差。

显然上述这些参数都是 $\Sigma = (\sigma_{ij})$ 的函数，为了给出它们的极大似然估计量，我们对样本协方差矩阵也作相应的分块。记

$$A = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})' = (a_{ij}),\ S = \frac{1}{n-1}A = (s_{ij})$$

以及

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{matrix} r \\ s \end{matrix},\ S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}\begin{matrix} r \\ s \end{matrix}$$

我们将称 $r_{ij} = s_{ij}/\sqrt{s_{ii}s_{jj}} = a_{ij}/\sqrt{a_{ii}\cdot a_{jj}}$

为 x_i 和 x_j 样本相关系数。

$$B = S_{12}S_{22}^{-1} = A_{12}A_{22}^{-1}$$

为样本回归系数矩阵

$$S_{1.2} = \frac{1}{n-1}A_{1.2} = (s_{ij.r+1,\cdots,p}) = \frac{1}{n-1}(a_{ij.r+1,\cdots,p})$$

为样本的偏(条件)协方差矩阵。

$$\begin{aligned} r_{ij.r+1,\cdots,p} &= s_{ij.r+1,\cdots,p}/\sqrt{s_{ii.r+1,\cdots,p}s_{jj.r+1,\cdots,p}} \\ &= a_{ij.r+1,\cdots,p}/\sqrt{a_{ii.r+1,\cdots,p}\cdot a_{jj.r+1,\cdots,p}} \end{aligned}$$

为 x_i 和 x_j 的样本偏相关系数。

若记

$$A=\begin{pmatrix} a_{11} & a'_{21} \\ a_{21} & A_{22} \end{pmatrix}\begin{matrix} 1 \\ p-1 \end{matrix} \qquad S=\begin{pmatrix} s_{11} & s'_{21} \\ s_{21} & S_{22} \end{pmatrix}\begin{matrix} 1 \\ p-1 \end{matrix}$$

我们称

$$r_{1.2,\cdots,p}=\sqrt{\frac{s'_{21}S_{22}^{-1}s_{21}}{s_{11}}}=\sqrt{\frac{a'_{21}A_{22}^{-1}a_{21}}{a_{11}}}$$

为 x_1 与 $x_2=(x_2,\cdots,x_p)'$ 之间的样本复相关系数。

下面引理 2.1 常称为统计方法中极大似然估计的不变性，它可用来从某些参数的极大似然估计推出它们的函数的极大似然估计。

引理 2.1 设一分布的参数 θ 的变化范围 Ω 是 R^k 中的一个区间，似然函数 $L(\theta)$ 是从 Ω 到实轴 R 的映象，设 $\hat{\theta}$ 是 θ 的极大似然估计，$\hat{\theta}$ 的值域是 Ω，$f(\theta)$ 是从 Ω 到 Ω^* 的一个变换，其中 Ω^* 是 $R^l(l\leqslant k)$ 中的一个区间，那么 $f(\theta)$ 的极大似然估计是 $f(\hat{\theta})$。

证明 对 Ω^* 中的任一元素 ω^*，考虑 Ω 中的集合

$$F(\omega^*)=\{\theta:\theta\in\Omega \text{ 且 } f(\theta)=\omega^*\}.$$

令

$$M(\omega^*)=\sup_{\theta\in F(\omega^*)}L(\theta)$$

显然 $M(\omega^*)$ 在 Ω^* 上有定义，它是由 $f(\theta)$ 导出的定义在 Ω^* 上的导出似然函数。对 Ω^* 中任二点 ω_1^* 和 ω_2^*，只要 $\omega_1^*\neq\omega_2^*$，则 $F(\omega_1^*)$ 和 $F(\omega_2^*)$ 没有共同的点，$\{F(\omega^*),\omega^*\in\Omega^*\}$ 构成了 Ω 的一个划分。$\hat{\theta}$ 一定属于且只属于某一个 $F(\omega^*)$，不妨设 $\hat{\omega}=f(\hat{\theta})$，于是 $\hat{\theta}\in F(\hat{\omega})$，$\hat{\omega}$ 是 Ω^* 中一个确定的元素，于是

$$L(\hat{\theta})=\sup_{\theta\in F(\hat{\omega})}L(\theta)=M(\hat{\omega})\geqslant L(\theta) \text{ 对一切 } \theta\in\Omega \text{ 成立.}$$

即

$$M(\hat{\omega})\geqslant\sup_{\theta\in F(\omega^*)}L(\theta)=M(\omega^*) \text{ 对一切 } \theta^*\in\Omega^* \text{ 成立.}$$

于是 $\hat{\omega}=f(\hat{\theta})$ 是 $\omega=f(\theta)$ 的极大似然估计。▌

定理 2.1 设 $x_1,\cdots,x_n$ 是从总体 $N_p(\mu,\Sigma)$ 抽取的容量为 n 的简单样本 $(\Sigma>0,n>p)$，那么

1）ρ_{ij} 的极大似然估计是 r_{ij}。

2）$\beta=\Sigma_{12}\Sigma_{22}^{-1}$ 和 $\Sigma_{1\cdot 2}$ 的极大似然估计是 $B=s_{12}S_{22}^{-1}$ 和 $\frac{1}{n}A_{1\cdot 2}$。

3）$\rho_{ij\cdot r+1,\cdots,p}$ 和 $\rho_{1\cdot 2,\cdots,p}$ 的极大似然估计是 $r_{ij\cdot r+1,\cdots,p}$，$r_{1\cdot 2,\cdots,p}$。

4）$|\Sigma|$的极大似然估计是 $\left|\frac{1}{n}A\right|$。

证明　由于 $\frac{1}{n}A$ 是 Σ 的极大似然估计，ρ_{ij}，β，$\Sigma_{1\cdot 2}$ 都是 σ_{ij} 的函数，而 $\rho_{ij\cdot r+1,\cdots,p}$ 由递推公式知道也是 σ_{ij} 的函数，$|\Sigma|$ 和 $\rho_{1\cdot 2,\cdots,p}$ 也同样是 σ_{ij} 的函数，从而由引理 2.1 知道它们相应的极大似然估计正是 $\frac{1}{n}A$ 所构成的相应统计量。 ▌

例 3.1　表 2.1 中列出了从 M、N 两地区各独立抽取的 30 块矿石样品，其中包含对六种化学元素的含量的测量数据。假设它们构成了一个来自 $N_6(\mu,\Sigma)$ 的样本，根据前面导出的有关公式，求出总体各参数的极大似然估计值。

解　a）M 地区

$$1)\ \hat{\boldsymbol{\mu}}_M=\bar{x}_M=\begin{pmatrix}401.4000\\37.5667\\27.6667\\38.3000\\32.8333\\32.1000\end{pmatrix}$$

$$2)\ \hat{\Sigma}_M=\frac{1}{30}A_M$$

$$=\begin{pmatrix}26315.9067 & 1372.6067 & 1283.0000 & -480.9867 & -286.8667 & 175.0933\\ & 455.9786 & 66.8222 & -27.8700 & -219.2056 & 184.8100\\ & & 223.4889 & 10.2667 & -37.1889 & 92.6667\\ & & & 85.6767 & 61.6167 & 1.0700\\ & & & & 213.1389 & -92.4500\\ & & & & & 190.0233\end{pmatrix}$$

3）样本相关矩阵 $R_M=(r_{ij})_M$

$$R_M=\begin{pmatrix}1 & 0.39625 & 0.52904 & -0.32033 & -0.12113 & 0.07830\\ & 1 & 0.20923 & -0.14100 & -0.70315 & 0.62784\\ & & 1 & 0.07419 & -0.17039 & 0.44967\\ & & & 1 & 0.45597 & 0.00839\\ & & & & 1 & -0.45938\\ & & & & & 1\end{pmatrix}$$

4）x_6 关于 $\boldsymbol{x}_1=(x_1,\cdots,x_5)'$ 的样本回归方程。
根据第二章第三节中记号知道若记

$$\boldsymbol{x}=\begin{pmatrix}\boldsymbol{x}_1\\ x_6\end{pmatrix},\ \boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}_1\\ \mu_6\end{pmatrix},\ \Sigma=\begin{pmatrix}\Sigma_{11} & \boldsymbol{\sigma}_{16}\\ \boldsymbol{\sigma}'_{16} & \sigma_{66}\end{pmatrix},$$

$$A=\begin{pmatrix}A_{11} & \boldsymbol{a}_{16}\\ \boldsymbol{a}'_{16} & a_{66}\end{pmatrix},\ \bar{\boldsymbol{x}}=\begin{pmatrix}\bar{\boldsymbol{x}}_1\\ \bar{x}_6\end{pmatrix}$$

则 x_6 关于 $\boldsymbol{x}_1=(x_1,\cdots,x_5)'$ 的总体回归方程为

$$x_6-\mu_6=\boldsymbol{\sigma}'_{16}\Sigma_{11}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1).$$

令 $\boldsymbol{\mu}$ 及 Σ 分别以它的极大似然估计 $\bar{\boldsymbol{x}}$ 和 $\frac{1}{n}A$ 代入得到 x_6 关于 $\boldsymbol{x}_1=(x_1,\cdots,x_5)'$ 的(样本)回归平面方程为

$$\begin{aligned}x_6=311.43056-0.86396x_1+1.14507x_2+0.79601x_3\\ -0.42641x_4+0.57118x_5\end{aligned}$$

5）x_6 关于 $(x_1,\cdots,x_5)$ 的样本复相关系数为

$$r_{6\cdot 1,\cdots,5}=0.86212$$

6）若干样本偏相关系数

$$r_{56\cdot 1}=-0.45463 \qquad r_{56\cdot 1,2}=-0.27326$$
$$r_{56\cdot 1,2,3}=0.18749$$

b）N 地区

1) $\hat{\mu}_N = \bar{x}_N = \begin{pmatrix} 355.1667 \\ 62.7333 \\ 12.4667 \\ 21.1667 \\ 8.2667 \\ 22.3667 \end{pmatrix}$

2) $\hat{\Sigma}_N = \frac{1}{30} A_N$

$$= \begin{pmatrix} 14586.4056 & 686.0111 & 262.5222 & -213.6944 & -21.2444 & -102.9278 \\ & 388.8622 & -22.5422 & -4.4556 & 6.1044 & 56.6311 \\ & & 40.4489 & -3.3444 & -1.5911 & 23.4289 \\ & & & 13.8056 & 1.1222 & 6.3389 \\ & & & & 1.5289 & 4.8022 \\ & & & & & 87.9656 \end{pmatrix}$$

3) 样本相关矩阵 $R_N = (r_{ij})_N$

$$R_N = \begin{pmatrix} 1 & 0.28804 & 0.34177 & -0.4760 & -0.14226 & -0.09087 \\ & 1 & -0.17974 & -0.06081 & 0.25036 & 0.30620 \\ & & 1 & -0.14153 & -0.20233 & 0.39277 \\ & & & 1 & 0.24427 & 0.18190 \\ & & & & 1 & 0.41409 \\ & & & & & 1 \end{pmatrix}$$

4) x_6 关于 $(x_1, \cdots, x_5)$ 的样本回归方程为

$$x_6 = 119.1611 - 0.38630x_1 + 0.44783x_2 + 0.68610x_3 + 0.02981x_4 + 0.37855x_5$$

5) x_6 对 $(x_1, \cdots, x_5)$ 的样本复相关系数

$$r_{6\cdot 1,\cdots,5} = 0.77710$$

6) 若干样本偏相关系数

$r_{56\cdot 1} = 0.40697$ $\quad$ $r_{56\cdot 1,2} = 0.33618$

$r_{56\cdot 1,2,3} = 0.49666$.

从上述计算的各项结果可以了解到 M, N 两地区矿石中六种

元素的平均含量，每个地区矿石中包含的元素含量的相互依赖关系，可以根据部分元素的含量推算出其他元素的含量等．这些都是十分有用的．另外我们往往还希望对两地区的矿石进行比较，例如希望了解这两地区矿石中元素含量除了随机性偏差外能否认为相同．如果不相同，那么那一个地区更具有利用价值等等．这些问题的解决将在第五章和以后各章节中逐步解决．

（二） 样本的几何解释

在本小段中我们将从几何角度来说明样本相关系数和样本广义方差的直观意义．设从 p 维总体获得一个容量为 n 的样本 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$，则样本矩阵为

$$X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)=\begin{pmatrix}\boldsymbol{x}^{(1)'}\\ \vdots\\ \boldsymbol{x}^{(p)'}\end{pmatrix}.$$

于是可以从两个角度来认识 X．一种观点是将 X 的列向量 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 看成是 R^p 中的 n 个点，那么 $\bar{\boldsymbol{x}}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i$ 正是这 n 个点的重心，这时

$$\operatorname{tr}A=\operatorname{tr}(X-\bar{\boldsymbol{x}}\mathbf{1}_n')(X-\bar{\boldsymbol{x}}\mathbf{1}_n')'=\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'(\boldsymbol{x}_i-\bar{\boldsymbol{x}})$$

是该 n 个点与重心之间的距离的平方和，它在一定程度上反映了总体协方差矩阵 Σ 的 $\operatorname{tr}\Sigma=\sum_{i=1}^{p}\sigma_{ii}$（总方差和的大小），从而从一个侧面反映了数据的变异性．

另一种观点是将 X 的各个行向量 $\boldsymbol{x}^{(i)}(i=1,\cdots,p)$ 看成是 R^n 中的 p 个点或者 p 个向量．而 $\mathbf{1}_n$ 作为 R^n 中的一向量，它与各坐标轴成相等的夹角，在该方向上的单位向量正是 $\frac{1}{\sqrt{n}}\mathbf{1}_n$．向量 $\boldsymbol{x}^{(i)}$ 在该方向上的投影正是

$$\left(\boldsymbol{x}^{(i)'}\frac{1}{\sqrt{n}}\mathbf{1}_n\right)\cdot\frac{1}{\sqrt{n}}\mathbf{1}_n=\frac{1}{n}\sum_{j=1}^{n}x_{ij}\mathbf{1}_n=\bar{x}_i\mathbf{1}_n$$

其中 $\bar{x}_i$ 是 $\bar{\boldsymbol{x}}$ 的第 i 个分量．因此 $\boldsymbol{x}^{(i)}-\bar{x}_i\mathbf{1}_n$ 将与该方向垂直，记

表 2.1

M 地区

样品号 \ 元素	x_1 Ba	x_2 Cr	x_3 Co	x_4 Pb	x_5 Cu	x_6 Ni
1	417	47	23	27	13	37
2	500	50	30	35	20	35
3	500	50	40	35	20	35
4	500	70	20	30	8	30
5	500	70	20	30	8	30
6	500	50	40	35	20	35
7	210	30	9	20	25	40
8	265	35	6	55	45	29
9	200	60	10	50	30	40
10	200	60	10	50	30	40
11	500	23	32	33	27	25
12	500	23	32	33	27	25
13	232	5	9	32	38	9
14	232	5	9	32	38	9
15	217	25	19	33	29	25
16	100	43	10	30	24	55
17	286	16	17	51	57	36
18	250	1	20	50	55	1
19	250	1	20	50	55	1
20	500	1	40	40	60	30
21	333	17	40	50	50	30
22	450	22	38	50	50	40
23	400	63	53	53	20	53
24	650	50	20	35	50	25
25	650	50	20	35	50	25
26	450	50	50	40	25	50
27	450	50	50	40	25	50
28	600	50	53	27	30	40
29	400	50	50	50	26	53
30	800	60	30	30	30	30

N 地 区 （续表）

元素 / 样品号	x_1 Ba	x_2 Cr	x_3 Co	x_4 Pb	x_5 Cu	x_6 Ni
1	200	50	10	25	7	20
2	200	43	11	23	6	11
3	450	60	10	20	8	20
4	200	50	10	25	7	20
5	200	50	10	25	7	20
6	333	83	13	22	9	23
7	150	41	10	20	9	22
8	150	41	10	20	9	22
9	450	60	10	20	8	20
10	450	60	10	20	8	20
11	283	73	11	25	7	23
12	283	80	12	23	9	22
13	350	86	11	24	10	34
14	150	41	10	20	9	22
15	300	100	10	25	10	30
16	490	57	34	16	7	39
17	500	60	10	15	7	1
18	500	60	10	15	7	1
19	300	6	11	23	10	25
20	350	85	10	18	10	30
21	400	80	10	23	10	18
22	490	57	34	16	7	39
23	500	60	10	15	7	1
24	550	73	11	17	7	27
25	425	75	11	24	8	28
26	350	85	10	18	10	30
27	400	80	10	23	10	28
28	451	31	25	30	9	15
29	450	80	10	20	8	20
30	250	75	10	25	8	30

$$e_i = x^{(i)} - \bar{x}_i \mathbf{1}_n$$

正是 $x^{(i)}$ 在投影后的剩余向量，若将 e_i 不改变大小和方向平移到以原点为向量起点，那么这个剩余向量长度的平方和正是

$$\|e_i\|^2 = e_i' e_i = (x^{(i)} - \bar{x}_i \mathbf{1}_n)'(x^{(i)} - \bar{x}_i \mathbf{1}_n)$$
$$= \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2$$

注意到第 i 个分量 x_i 的样本方差正是

$$s_{ii} = \frac{1}{n-1} a_{ii} = \frac{1}{n-1} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2$$
$$= \frac{1}{n-1} \|e_i\|^2 \tag{2.2}$$

于是 e_i 的长度与第 i 分量的样本均方差成比例。

利用向量的内积知道：设剩余向量 e_i 和 e_k 的夹角为 θ_{ik}，那么

$$\langle e_i, e_k \rangle = e_i' e_k = \|e_i\| \|e_k\| \cos\theta_{ik}$$
$$= \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)$$

所以

$$\cos\theta_{ik} = \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k) \Big/ \sqrt{\sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 \cdot \sum_{j=1}^{n} (x_{kj} - \bar{x}_k)^2}$$
$$= s_{ik} / \sqrt{s_{ii} \cdot s_{kk}}$$
$$= r_{ik} \tag{2.3}$$

这说明分量 x_i 和 x_k 的样本相关系数正是剩余向量 e_i 和 e_k 的夹角余弦的大小。于是如果剩余向量几乎同向，样本相关系数接近于 1，如果垂直，则样本相关系数几乎为 0，如果剩余向量几乎反向，则样本相关系数接近 -1。

样本广义方差 $|S|$ 提供了把所有样本方差和协方差合并为一个数量指标的方法。作为 $|\Sigma|$ 的估计是有用的。下面给出 $|S|$ 的一个直观解释，首先考虑 $p=2$ 时，两个剩余向量 $\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$ 所生成的平行四边形的面积的绝对值为

$$\begin{aligned} S_{\square} &= \|\boldsymbol{e}_1\| \|\boldsymbol{e}_2\| |\sin\theta_{12}| \\ &= (n-1)\sqrt{s_{11}s_{22}\cdot(1-r_{12}^2)} \end{aligned}$$

其中第二式是利用了(2.2)和(2.3)式。

另一方面有

$$|S| = \begin{vmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{vmatrix} = s_{11}s_{22} - s_{12}s_{21} = s_{11}s_{22}(1-r_{12})^2$$

从而

$$|S| = S_{\square}^2/(n-1)^2 \tag{2.4}$$

由归纳法可以证明（参考 [7]）；由 p 个剩余向量 $\boldsymbol{e}_i$ 所张成的(广义)平行六面体的体积与广义方差之间有关系式

$$|S| = (\text{体积})^2/(n-1)^p \tag{2.5}$$

由此可见，当剩余向量 $\boldsymbol{e}_i$ 的长度增加时（即 s_{ii} 增加时）体积增大，$|S|$变大。或者 $\boldsymbol{e}_i$ 与 $\boldsymbol{e}_k$ 的夹角趋于垂直(即相关系数 r_{ik} 趋于 0 时)，体积增大，$|S|$ 变大。由于 $\boldsymbol{e}_i$ 反映了第 i 个分量的样本值偏离它的平均值的大小，那么由 $\boldsymbol{e}_i$ 张成的体积可近似地看成 p 个分量偏离平均向量的一个综合度量。当然以一个数度量 p 个变量的偏差的综合状态总是有不足之处。例如当恰好有一个 s_{ii} 很小或者有一剩余向量几乎处于其他二个或者二个以上剩余向量张成的超平面上时，都将使得它的体积近似于 0。

定理 2.2 样本广义方差为 0 的充分必要条件是至少有一个剩余向量在其他剩余向量的线性组合所形成的超平面上。

证明 由于

$$E = \begin{pmatrix} \boldsymbol{e}_p' \\ \vdots \\ \boldsymbol{e}_1' \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}^{(1)'} - x_1\boldsymbol{1}_n' \\ \vdots \\ \boldsymbol{x}^{(p)'} - x_p\boldsymbol{1}_n' \end{pmatrix} = X - \bar{\boldsymbol{x}}\boldsymbol{1}_n'$$

从而 $S=\frac{1}{n-1}E\cdot E'$. 由第一章定理1.4知道 $\mathrm{rk}(S)=\mathrm{rk}(E)$. 而 $|S|=0$ 等价于 $\mathrm{rk}(S)<p$. 即 $\mathrm{rk}(E)<p$. 它的充分必要条件正是诸 e_i 中存在线性相关. 定理成立. ▌

（三） 样本广义方差及样本相关系数的分布

设 $x_1,\cdots,x_n$ 相互独立且服从 $N_p(\mu,\Sigma)(n>p,\Sigma>0)$. 样本的协方差矩阵为

$$S=\frac{1}{n-1}A,\qquad \left(A=\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'\right)$$

从而样本广义方差

$$|S|=(n-1)^{-p}|A|$$

由第三章定理2.3知 $A\sim W_p(n-1,\Sigma)$. 又由第三章推论2.7有

$$|S|=(n-1)^{-p}\cdot|\Sigma|\cdot\prod_{i=1}^{p}u_i$$

其中 u_i 相互独立且分别服从 $\chi^2(n-i)(i=1,\cdots,p)$.
因此当 $p=1$ 时 $|S|$ 的分布正是 $c\chi^2(n-1)$. 当 $p=2$ 时可以证明(参考文献[15])

$$2(n-1)\cdot\sqrt{|S|/|\Sigma|}\sim\chi^2(2n-2).$$

当 $p\geqslant 3$ 时很难求出它的精确分布, Anderson (参考文献 [7]) 指出: 当 $n\to\infty$ 时

$$\sqrt{n-1}\left(\frac{|S|}{|\Sigma|}-1\right)\text{ 渐近服从 }N(0,2p).$$

设总体服从 $N_p(\mu,\Sigma)$. $\Sigma>0,n>p,\rho_{ij}=\sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ 是总体相关系数, $r_{ij}=a_{ij}/\sqrt{a_{ii}a_{jj}}$ 是从 n 个样品求得的样本相关系数. 它的分布与 ρ_{ij} 有着密切的关系. 当 $p=2$ 时,我们可以求得样本相关系数 r 的分布,而当 $p>2$ 时,仅当 Σ 是对角阵时, 我们能够求出 $R=(r_{ij})$ 样本相关矩阵的分布. 下面分别对 $p=2$ 和 $p>2$ 进行讨论.

1) 当 $p=2$

设总体 $x \sim N_2(\boldsymbol{\mu},\Sigma)$ 其中

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} > 0$$

从该总体抽取的样本矩阵为 $X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)(n>2)$ 则 $X \sim N_{2,n}(\boldsymbol{\mu}\mathbf{1}_n',\Sigma,I_n)$ 此时

$$\begin{aligned} A &= XX' - n\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}' \\ &= \sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})' \sim W_2(n-1,\Sigma). \end{aligned}$$

记

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

样本相关系数为 $r = a_{12}/\sqrt{a_{11}\cdot a_{22}}$.

首先我们作一简化，令 $\tilde{A} = DAD, D=\operatorname{diag}(\sigma_1^{-1},\sigma_2^{-1})$，则

$$\tilde{A} = \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{12} \\ \tilde{a}_{21} & \tilde{a}_{22} \end{pmatrix} \sim W_2(n-1,\tilde{\Sigma}),\ \tilde{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} > 0$$

而且

$$\tilde{r} = \tilde{a}_{12}/\sqrt{\tilde{a}_{11}\tilde{a}_{22}} = a_{12}/\sqrt{a_{11}a_{22}} = r.$$

因此要求 r 的分布，不失一般性可设 $\Sigma=\tilde{\Sigma}$. 我们分 $\rho=0$ 和 $\rho\neq 0$ 两种情况进行讨论.

i) $\rho=0$ 的情况

定理 2.3 当总体相关系数 $\rho=0$ 时，二维正态总体 $N_2(\boldsymbol{\mu},\Sigma)$ $(\Sigma>0)$ 的样本相关系数 r 的密度为

$$f(r) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}(1-r^2)^{\frac{n-4}{2}} \tag{2.6}$$

证明 由 $\rho=0$，知道 $A \sim W_2(n-1,I_2)$. 于是 A 的密度也即 a_{11},a_{22},a_{12} 的联合密度为

$$c(2,n-1)\cdot(a_{11}a_{22}-a_{12}^2)^{\frac{n-4}{2}}\exp\left[-\frac{1}{2}(a_{11}+a_{22})\right]$$

作变换 $a_{12}=r\sqrt{a_{11}a_{22}}$,则

$$J(a_{11},a_{22},a_{12}\to a_{11},a_{22},r)=\sqrt{a_{11}a_{22}}.$$

于是 (a_{11},a_{22},r) 的联合密度为

$$c(2,n-1)\cdot(1-r^2)^{\frac{n-4}{2}}(a_{11}a_{22})^{\frac{1}{2}(n-1)-1}\exp\left(-\frac{a_{11}+a_{22}}{2}\right) \tag{2.7}$$

关于 a_{11},a_{22} 分别积分。由

$$\int_0^\infty a_{ii}^{\frac{1}{2}(n-1)-1}\exp\left(-\frac{a_{ii}}{2}\right)da_{ii}=2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right) \tag{2.8}$$

以及

$$c(2,n-1)=\left[2^{n-1}\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)\right]^{-1}$$

代入(2.7)式可直接求出 r 有密度(2.6)式,定理得证。▮

定理 2.4 当总体相关系数 $\rho=0$ 时，设 r 是二维正态总体 $N_2(\boldsymbol{\mu},\Sigma)(\Sigma>0)$ 的样本相关系数,则

$$T=\sqrt{n-2}\cdot r/\sqrt{1-r^2}$$

服从自由度为 $n-2$ 的 t 分布

证明 在 r 的密度(2.6)式中作变换 $t=\sqrt{n-2}\,r/\sqrt{1-r^2}$，于是

$$(1-r^2)=\left(1+\frac{t^2}{n-2}\right)^{-1},$$

$$J(r\to t)=\frac{1}{\sqrt{n-2}}\left(1+\frac{t^2}{n-2}\right)^{-\frac{3}{2}}$$

从而 T 的密度为:

$$f(t)=\frac{1}{\sqrt{(n-2)\pi}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}\left(1+\frac{t^2}{n-2}\right)^{-\frac{1}{2}(n-1)}$$

正是自由度为 $n-2$ 的 t 分布. ∎

在实用中为了检验二个随机变数是否不相关（对正态总体即检验它们是否独立），只须检验它们的相关系数是否为 0. 因此通常先作假设 $H_0: \rho=0$，根据试验数据计算出 r，看 $|r|$ 是否太大，当 $|r|$ 大到使得 $\sqrt{n-2}\,|r|/\sqrt{1-r^2}$ 大于 $t_{1-\frac{\alpha}{2}}(n-2)$ 时（α 是给定的显著水平）时我们拒绝 $H_0: \rho=0$. 否则我们可以接受 $\rho=0$. 具体的细节将在第五章第五节关于独立性检验中进行讨论.

ii) $\rho\neq 0$ 的情况

定理 2.5 当总体相关系数 $\rho\neq 0$ 时，二维正态总体 $N_2(\boldsymbol{\mu},\Sigma)$ $(\Sigma>0)$ 的样本相关系数 r 的密度为

$$f(r)=2^{n-3}\cdot[\pi\Gamma(n-2)]^{-1}(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}$$
$$\cdot\sum_{j=0}^{\infty}\Gamma^2\left(\frac{n-1+j}{2}\right)\frac{(2r\rho)^j}{j!} \tag{2.9}$$

证明见本章附录.

由于 $\rho\neq 0$ 时，二维正态总体样本相关系数 r 的分布过于复杂，我们常利用下面的渐近分布.

设

$$z=\frac{1}{2}\ln\frac{1+r}{1-r},\qquad \zeta=\frac{1}{2}\ln\frac{1+\rho}{1-\rho}$$

那么当 $n\to\infty$ 时 $\sqrt{n-2}\,(z-\zeta)$ 渐近服从 $N(0,1)$. 即当 n 充分大时 $z\sim N\left(\zeta,\frac{1}{n-2}\right)$ 渐近成立.

上述定理的结论可推广到 p 维正态总体的样本相关系数 r_{ij} 的分布，即

推论 2.1 设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 是从总体 $N_p(\boldsymbol{\mu},\Sigma)$ 中抽取的容量为 n 的简单样本（$n>p$, $\Sigma>0$），ρ_{ij} 是总体中分量 x_i 和 x_j 的相关系数，r_{ij} 是相应的样本相关系数，那末 r_{ij} 的分布密度是在

(2.9)式中用 r_{ij} 代替 r，ρ_{ij} 代替 ρ 后得到的密度。

证明　设

$$A = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})' \left(\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \right),$$

由第三章定理(2.3)知 $A \sim W_p(n-1, \Sigma)$. $\Sigma = (\sigma_{ij})$, 又由第三章定理 2.5 知道

$$\begin{pmatrix} a_{ii} & a_{ji} \\ a_{ij} & a_{jj} \end{pmatrix} \sim W_2\left(n-1, \begin{pmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ji} & \sigma_{jj} \end{pmatrix}\right)$$

再利用定理 2.5 推论即可得证. ▮

下面推论 2.2 将利用 $p = 2$ 的结论推出 p 维正态总体的样本偏相关系数的分布。

设 $A \sim W_p(n-1, \Sigma)$, $\Sigma > 0$, $n > p$. 且将 Σ, A 作下列划分

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}_s^r, \qquad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}_s^r,$$

其中 A_{11} 是 $r \times r$ 阶, Σ_{22} 是 $s \times s$ 阶, $r + s = p$, A 有相应的划分，此时总体偏协方差矩阵和样本偏协方差矩阵分别为

$$\Sigma_{1\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = (\sigma_{ij\cdot r+1,\cdots,p})$$

$$A_{1\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21} = (a_{ij\cdot r+1,\cdots,p})$$

总体偏相关系数为

$$\rho_{ij\cdot r+1,\cdots,p} = \sigma_{ij\cdot r+1,\cdots,p} / \sqrt{\sigma_{ii\cdot r+1,\cdots,p} \cdot \sigma_{jj\cdot r+1,\cdots,p}}$$

样本偏相关系数为

$$r_{ij\cdot r+1,\cdots,p} = a_{ij\cdot r+1,\cdots,p} / \sqrt{a_{ii\cdot r+1,\cdots,p} \cdot a_{jj\cdot r+1,\cdots,p}}.$$

推论 2.2　如上所设，那么变量 x_i 和 x_j 的样本偏相关系数 $r_{ij\cdot r+1,\cdots,p}$ 的密度正是在(2.9)式中以 $r_{ij\cdot r+1,\cdots,p}$, $\rho_{ij\cdot r+1,\cdots,p}$ 和 $n-p+r-1$ 代替 r, ρ 和 $n-1$ 所得的密度。

证明　由第三章定理 2.6 知道

$$A_{1\cdot 2} \sim W_r(n-p+r-1, \Sigma_{1\cdot 2})$$

利用推论 2.1 的证明立即可得证。▌

2）当 $p \geqslant 3$.

设 $X = (x_1, \cdots, x_n) \sim N_{p,n}(\mu 1_n', \Sigma, I_n)$ $(\Sigma > 0, n > p)$ 其中

$$\Sigma = \operatorname{diag}(\sigma_{11}, \cdots, \sigma_{pp}) \text{ 是一对角阵}$$

此时 $A = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})' = (a_{ij}) \sim W_p(n-1, \Sigma)$

记相应的样本相关系数矩阵为 $R = (r_{ij})$，那末 R 的分布(即 r_{ij} 的联合分布)有下列结论。

定理 2.6 如上所设，则样本相关系数矩阵 R 的密度为

$$\left[\Gamma\left(\frac{n-1}{2}\right)\right]^p \left[\Gamma_p\left(\frac{n-1}{2}\right)\right]^{-1} |R|^{\frac{1}{2}(n-p-2)} \qquad (2.10)$$

且 R 与 $a_{11}, \cdots, a_{pp}$ 相互独立。

证明：令 $\tilde{A} = DAD$. $D = \operatorname{diag}(\sigma_{11}^{-\frac{1}{2}}, \cdots, \sigma_{pp}^{-\frac{1}{2}})$. 则

$$\tilde{A} = (\tilde{a}_{ij}) \sim W_p(n-1, I_p).$$

而此时

$$\tilde{r}_{ij} = \tilde{a}_{ij}/\sqrt{\tilde{a}_{ii}\tilde{a}_{jj}} = a_{ij}/\sqrt{a_{ii}a_{jj}} = r_{ij}.$$

因此不失一般性，可设 $A \sim W_p(n-1, I_p)$. 它的密度为

$$c(p, n-1) \cdot |A|^{\frac{n-p-2}{2}} \operatorname{etr}\left(-\frac{1}{2}A\right)$$

作变换

$$\begin{cases} a_{ii} = a_{ii} & (i = 1, \cdots, p) \\ a_{ij} = r_{ij}\sqrt{a_{ii}a_{jj}} & (1 \leqslant i < j \leqslant p) \end{cases}$$

此时

$$\begin{aligned} J(A \to a_{11}, \cdots, a_{pp}, R) &= J(a_{ij}(1 \leqslant i \leqslant j \leqslant p) \\ &\to a_{11}, \cdots, a_{pp}, r_{ij}, 1 \leqslant i < j \leqslant p) \\ &= \prod_{i<j} \sqrt{a_{ii}a_{jj}} = \prod_{i=1}^{p} (a_{ii})^{\frac{p-1}{2}} \end{aligned}$$

因此R与 $a_{11},\cdots,a_{pp}$ 的联合密度为

$$c(p,n-1)\cdot\prod_{i=1}^{p}a_{ii}^{\frac{n-3}{2}}|R|^{\frac{1}{2}(n-p-2)}\exp\left(-\frac{1}{2}\sum_{i=1}^{p}a_{ii}\right)$$

由此可知R与 $a_{11},\cdots,a_{pp}$ 相互独立，关于 a_{ii} 积分，由于

$$\int_0^\infty a_{ii}^{\frac{n-3}{2}}\exp\left(-\frac{1}{2}a_{ii}\right)da_{ii}=2^{\frac{1}{2}(n-1)}\Gamma\left(\frac{n-1}{2}\right)$$

计算后即得R的密度为(2.10)式. ▌

当 $\Sigma=\text{diag}(\sigma_{11},\cdots,\sigma_{pp})$ 时，r_{ij} 的边缘分布正是定理 2.3 中所给出的(2.6)式，只须将 r_{ij} 代替(2.6)式中的 r 即可.

最后我们将给出样本复相关系数 $r_{1\cdot 2,\cdots,p}$ 的分布.

设总体 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\Sigma)$ $\Sigma>0$，$X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)(n>p)$是从该总体抽取的样本矩阵，记 $A=\sum\limits_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'$. 则 $A\sim W_p(n-1,\Sigma)$.

为了表达方便，记

$$\boldsymbol{x}=\begin{pmatrix}x_1\\ \boldsymbol{x}_2\end{pmatrix}\begin{matrix}1\\ p-1\end{matrix}\quad \Sigma=\begin{pmatrix}\sigma_{11} & \boldsymbol{\sigma}'_{21}\\ \boldsymbol{\sigma}_{21} & \Sigma_{22}\end{pmatrix}\quad A=\begin{pmatrix}a_{11} & \boldsymbol{a}'_{21}\\ \boldsymbol{a}_{21} & A_{22}\end{pmatrix}$$

其中 $\boldsymbol{x}_2=(x_2,\cdots,x_p)'$. Σ_{22},A_{22} 都是$(p-1)\times(p-1)$阶的.

总体复相关系数为

$$\rho_{1\cdot 2,\cdots,p}=\sqrt{\boldsymbol{\sigma}'_{21}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}/\sigma_{11}}$$

它的极大似然估计正是样本复相关系数

$$r_{1\cdot 2,\cdots,p}=\sqrt{\boldsymbol{a}'_{21}A_{22}^{-1}\boldsymbol{a}_{21}/a_{11}}$$

下面将简记 $\tilde{\rho}=\rho_{1\cdot 2,\cdots,p}$ 和 $\tilde{r}=r_{1\cdot 2,\cdots,p}$

i) 当 $\tilde{\rho}=0$ 时 $\tilde{r}$ 的分布

定理 2.7 当总体复相关系数 $\tilde{\rho}=0$ 时，样本复相关系数 $\tilde{r}$ 有下列性质

$$F=\frac{n-p}{p-1}\frac{\tilde{r}^2}{1-\tilde{r}^2}\sim F(p-1,n-p) \tag{2.11}$$

证明 首先注意到 $\bar{\rho}=0$ 等价于 $\beta=\sigma'_{21}\Sigma_{22}^{-1}=0$，即 $\sigma_{1\cdot 2}=\sigma_{11}$. 因此由第三章定理 2.6 知道

$$a_{1\cdot 2}=a_{11}-\alpha'_{21}A_{22}^{-1}\alpha_{21}\sim W_1(n-p,\sigma_{11}) \text{ 即 } \sigma_{11}\chi^2(n-p).$$

且与 (A_{22},α_{21}) 相互独立，而当 A_{22} 给定时，

$$\alpha_{21}\sim N_{p-1}(0,\sigma_{11}A_{22}).$$

由第三章定理 1.2 知 $\alpha'_{21}A_{22}^{-1}\alpha_{21}$ 的条件分布服从 $\sigma_{11}\chi^2(p-1)$ 且与 A_{22} 无关，从而 $\alpha'_{21}A_{22}^{-1}\alpha_{21}$ 的无条件分布仍是 $\sigma_{11}\chi^2(p-1)$. 并与 $a_{1\cdot 2}$ 相互独立，由于 $1-\bar{r}^2=a_{1\cdot 2}/a_{11}$，所以

$$\frac{n-p}{p-1}\frac{\bar{r}^2}{1-\bar{r}^2}=\frac{n-p}{p-1}\cdot\frac{a'_{21}A_{22}^{-1}a_{21}/\sigma_{11}}{a_{1\cdot 2}/\sigma_{11}}$$

根据 F 分布定义，知道 $F\sim F(p-1,n-p)$. ▮

注意到当 $p=2$ 时复相关系数 $\bar{r}$ 正是普通的相关系数 r，定理 2.4 $\sqrt{n-2}\cdot r/\sqrt{1-r^2}\sim t(n-2)$ 正与本定理的结论相吻合.

由 $F=\dfrac{n-p}{p-1}\cdot\bar{r}^2/1-\bar{r}^2\sim F(p-1,n-p)$，直接可以从 F 分布密度中作适当变换得出 $\bar{r}$ 的分布密度，这一工作留待读者自行完成.

ii) 当 $\bar{\rho}\neq 0$ 时 $\bar{r}$ 的分布

定理 2.8 当总体复相关系数 $\bar{\rho}\neq 0$ 时，则样本复相关系数 $\bar{r}$ 有下列密度

$$f(\bar{r})=\frac{2(1-\bar{\rho}^2)^{\frac{n-1}{2}}(1-\bar{r}^2)^{\frac{n-p-2}{2}}}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-p}{2}\right)}$$

$$\cdot\sum_{k=0}^{\infty}\frac{\Gamma^2\left(\frac{n-1}{2}+k\right)}{k!\,\Gamma\left(\frac{p-1}{2}+k\right)}(\bar{\rho}^2)^k(\bar{r}^2)^{\frac{p}{2}+k-1} \tag{2.12}$$

证明 见本章附录.

习 题 四

4.1 设 $x_i\ (i=1,\cdots,n)$ 是从总体 $N_p(\mu,\Sigma)$ 中独立抽取的容量为 n 的简单样本，$\Sigma>0$，$n>p$。记

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\quad A=\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'.$$

x 是抽自该总体的另一个独立随机样本，证明

1) $x-\bar{x}\sim N_p\left(0,\left(1+\frac{1}{n}\right)\Sigma\right)$.

2) $\frac{n-p}{p}\cdot\frac{n}{n+1}(x-\bar{x})'A^{-1}(x-\bar{x})\sim F(p,n-p)$.

4.2 设 $x_i(i=1,\cdots,n)$ 是从总体 $N_p(\mu,\sigma^2I_p)$ 中独立抽取的容量为 n 的简单样本，$n>p$，试求 μ 及 σ^2 的极大似然估计，

4.3 设 y_i 是从参数为 μ 及 Σ 的 p 维对数正态分布总体中独立抽取的容量为 n 的简单样本（见习题二第 16 题）证明 μ 和 Σ 的极大似然估计与正态总体完全一样即

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}\ln y_i\triangleq\ln\bar{y}$$

$$\hat{\Sigma}=\frac{1}{n}\sum_{i=1}^{n}(\ln y_i-\ln\bar{y})(\ln y_i-\ln\bar{y})'$$

4.4 设 $x_i(i=1,\cdots,n_1)$ 是从正态总体 $N_p(\mu,\Sigma)$ 中独立抽取的容量为 n_1 的简单样本，而 $y_i(i=1,\cdots,n_2)$ 是从正态总体 $N_p(\nu,\Sigma)$ 中独立抽取的容量为 n_2 的简单样本（$\Sigma>0$）且两样本是相互独立的，记

$$\bar{x}=\frac{1}{n_1}\sum_{i=1}^{n_1}x_i,\quad \bar{y}=\frac{1}{n_2}\sum_{i=1}^{n_2}y_i.$$

$$A_1=\sum_{i=1}^{n_1}(x_i-\bar{x})(x_i-\bar{x})',\ A_2=\sum_{i=1}^{n_2}(y_i-\bar{y})(y_i-\bar{y})'.$$

证明

1) μ,ν,Σ 的极大似然估计分别为

$$\hat{\mu}=\bar{x},\ \hat{\nu}=\bar{y},\ \hat{\Sigma}=\frac{1}{n_1+n_2}(A_1+A_2)$$

2) 证明 $\hat{\mu},\hat{\nu},\hat{\Sigma}$ 相互独立，并指出它们的分布。

*4.5 设 $x_i(i=1,\cdots,n)$ 是从 $N_2(\mu,\Sigma)$ 中独立抽取的容量为 $n(n>2)$ 的简单样本，其中 $\Sigma=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$，$|\rho|<1$，$\rho$ 是总体相关系数，记

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\quad A=\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'=\begin{pmatrix}a_{11} & a_{21}\\ a_{12} & a_{22}\end{pmatrix}$$

则 $A\sim W_2(n-1,\Sigma)$。$r=a_{12}/\sqrt{a_{11}a_{22}}$ 是样本相关系数，证明

$$\frac{r}{\sqrt{1-r^2}}=\frac{1}{t_{22}}\left(\frac{\rho}{\sqrt{1-\rho^2}}t_{11}+t_{12}\right)$$

其中 $t_{11}^2\sim\chi^2(n-1)$，$t_{22}^2\sim\chi^2(n-2)$，$t_{12}\sim N(0,1)$ 且相互独立。

提示 令 $\Sigma=C\cdot C'$，其中

$$C=\begin{pmatrix}1 & 0\\ \rho & \sqrt{1-\rho^2}\end{pmatrix}$$

于是 $CAC'=\tilde{A}\sim W_2(n-1,I_2)$。利用第三章定理 2.14 的 Barlett 三角分解 $\tilde{A}=TT'$ 即可证明。

4.6 设 $x_i(i=1,\cdots,n)$ 是从总体 $N_p(\mu,\Sigma)$ 中抽取的容量为 $n(n>p)$ 的简单样本，试用微分方法证明：μ 和 Σ 的极大似然估计分别为 $\bar{x}$ 和 $\frac{1}{n}\Sigma$。其中 $\bar{x}$ 和 A 是习题 4.5 中所定义的。

提示：利用第一章例 4.2 和例 4.4 的结论。

附 录

(一) 定理 2.5 的证明

如同定理 2.4 以前的说明，令

$$\Sigma=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix},\ A=\begin{pmatrix}a_{11} & a_{21}\\ a_{12} & a_{22}\end{pmatrix},$$

于是

$$A\sim W_2(n-1,\Sigma),\ |\Sigma|=1-\rho^2,$$

$$\Sigma^{-1}=\frac{1}{1-\rho^2}\begin{pmatrix}1 & -\rho\\ -\rho & 1\end{pmatrix},\ r=a_{12}/\sqrt{a_{11}a_{22}}$$

以及

$$\mathrm{tr}(\Sigma^{-1}A)=\frac{1}{1-\rho^2}(a_{11}-2\rho a_{12}+a_{22})$$

从而 a_{11},a_{12},a_{22} 的联合密度为

$$c(2, n-1)|\Sigma|^{-\frac{n-1}{2}}|A|^{\frac{n-4}{2}}\operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}A\right)$$

$$= c(2, n-1)\cdot(1-\rho^2)^{-\frac{n-1}{2}}(a_{11}a_{22}-a_{12}^2)^{\frac{n-4}{2}}$$

$$\cdot\exp\left[-\frac{1}{1-\rho^2}(a_{11}+a_{22}-2\rho a_{12})\right]$$

作变换

$$\begin{cases} a_{11}=(1-\rho^2)v_{11} \\ a_{22}=(1-\rho^2)v_{22} \\ a_{12}=(1-\rho^2)r\sqrt{v_{11}v_{22}} \end{cases}$$

此时 $J(a_{11}, a_{12}, a_{22}\to r, v_{11}, v_{22})=(1-\rho^2)^3\cdot\sqrt{v_{11}v_{22}}$

由此可得 (r, v_{11}, v_{22}) 的联合密度为

$$f(r, v_{11}, v_{22})=c(2, n-1)\cdot(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}$$

$$\cdot(v_{11}v_{22})^{\frac{n-3}{2}}\exp\left\{-\frac{1}{2}(v_{11}+v_{22})+r\rho v_{11}^{\frac{1}{2}}v_{22}^{\frac{1}{2}}\right\} \quad (1)$$

为了求 r 的密度，只须对 v_{11}, v_{22} 进行积分，为此进行幂级数展开：

$$\exp(r\rho v_{11}^{\frac{1}{2}}v_{22}^{\frac{1}{2}})=\sum_{j=0}^{\infty}\frac{1}{j!}(r\rho v_{11}^{\frac{1}{2}}v_{22}^{\frac{1}{2}})^j$$

记

$$f_0(r)=c(2, n-1)(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}$$

(1)式可写成

$$f(r, v_{11}, v_{22})=f_0(r)\cdot\sum_{j=0}^{\infty}\frac{1}{j!}(\rho r)^j v_{11}^{\frac{1}{2}(n-3+j)}e^{-\frac{1}{2}v_{11}}\cdot v_{22}^{\frac{1}{2}(n-3+j)}e^{-\frac{1}{2}v_{22}}$$

而

$$\int_0^{\infty}v_{ii}^{\frac{1}{2}(n-3+j)}e^{-\frac{1}{2}v_{ii}}\,dv_{ii}=2^{\frac{1}{2}(n-1+j)}\Gamma\left(\frac{n-1+j}{2}\right)\quad(i=1, 2)$$

另一方面利用 $c(2, n-1)$ 表达式和 Γ 函数的倍值公式，可以求得 r 的密度为

$$f_0(r)\cdot\sum_{j=0}^{\infty}\frac{(r\rho)^j}{j!}\cdot 2^{n-1+j}\cdot\Gamma^2\left(\frac{n-1+j}{2}\right)$$

$$=2^{n-3}[\pi\Gamma(n-2)]^{-1}\cdot(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{1}{2}(n-4)}$$

$$\cdot\sum_{j=0}^{\infty}\Gamma^2\left(\frac{n-1+j}{2}\right)\cdot\frac{(2r\rho)^j}{j!}. \quad ∎$$

（二）定理 2.8 的证明

为了求出 $\bar{\rho}\neq 0$ 时 r 的分布，我们分二步进行

1）设 $A\sim W_p(n-1,\Sigma)$ $(n>p,\Sigma>0)$ 其中

$$\Sigma=\begin{pmatrix}\sigma_{11} & \sigma_{21}' \\ \sigma_{21} & \Sigma_{22}\end{pmatrix},\ A=\begin{pmatrix}a_{11} & a_{21}' \\ a_{21} & A_{22}\end{pmatrix}$$

当总体复相关系数 $\bar{\rho}\neq 0$ 时，下面证明：当 A_{22} 给定时

$$\frac{n-p}{p-1}\frac{r^2}{1-r^2}\Big| A_{22}\sim F(p-1,n-p,\lambda)$$

这是因为当 A_{22} 给定后 按第三章定理 2.6 知道

$$A_{1\cdot 2}=a_{11}-a_{21}'A_{22}^{-1}a_{21}\sim\sigma_{1\cdot 2}\chi^2(n-p)$$

且与 (a_{21},A_{22}) 相互独立，其中 $\sigma_{1\cdot 2}=\sigma_{11}-\sigma_{21}'\Sigma_{22}^{-1}\sigma_{21}$.

而
$$a_{21}\sim N_{p-1}(A_{22}\beta,\sigma_{1\cdot 2}A_{22})\ (\beta=\Sigma_{22}^{-1}\cdot\sigma_{21})$$

从而

$$a_{21}'A_{22}^{-1}a_{21}/\sigma_{1\cdot 2}\sim\chi^2(p-1,\lambda),$$

$(\lambda=\beta'A_{22}[A_{22}\cdot\sigma_{1\cdot 2}]^{-1}A_{22}\beta=\beta'A_{22}\beta/\sigma_{1\cdot 2})$ 且与 $a_{1\cdot 2}$ 独立

由此

$$\frac{n-p}{p-1}\frac{r^2}{1-r^2}=\frac{n-p}{p-1}\frac{a_{21}'A_{22}^{-1}a_{21}}{a_{1\cdot 2}}\sim F(p-1,n-p,\lambda).$$

2）由于 $\sigma_{1\cdot 2}=\sigma_{11}-\sigma_{21}'A_{22}^{-1}\sigma_{21}=\sigma_{11}(1-\bar{\rho}^2)$，$A_{22}\sim W_{p-1}(n-1,\Sigma_{22})$

所以

$\beta'A_{22}\beta\sim W_1(n-1,\beta'\Sigma_{22}\beta)$ 即 $\beta'A_{22}\beta\sim(\beta'\Sigma_{22}\beta)\cdot\chi^2(n-1)$.

由于 $\beta'\Sigma_{22}\beta=\sigma_{21}'\Sigma_{22}^{-1}\sigma_{21}=\sigma_{11}\bar{\rho}^2$. 于是

$$\lambda=\frac{\beta'A_{22}\beta}{\sigma_{1\cdot 2}}\sim\frac{\bar{\rho}^2}{1-\bar{\rho}^2}\cdot\chi^2(n-1)$$

记 $\nu=\frac{1-\bar{\rho}^2}{\bar{\rho}^2}$，上式可改为 $\nu\cdot\lambda\sim\chi^2(n-1)$. 由 $\chi^2(n-1)$ 密度可知 λ 的密度为

$$g(\lambda)=\frac{1}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{\nu}{2}\right)^{\frac{n-1}{2}}\lambda^{\frac{n-1}{2}-1}\exp\left(-\frac{1}{2}\nu\lambda\right)\tag{2}$$

由1)知道，当 A_{22} 给定(即 λ 给定时)

$$t=\frac{n-p}{p-1}\frac{r^2}{1-r^2}\sim F(p-1,n-p,\lambda).$$

它有密度(见第三章(1.9)式)

$$f(t)=e^{-\frac{\lambda}{2}}\sum_{i=0}^{\infty}\frac{1}{i!}\left(\frac{\lambda}{2}\right)^{i}(n-p)^{\frac{n-p}{2}}(p-1)^{\frac{p-1}{2}+i}$$

$$\cdot c_i\frac{t^{\frac{p-1}{2}-1+i}}{[(n-p)+(p-1)t]^{\frac{n-1}{2}+i}} \tag{3}$$

其中

$$c_i=\Gamma\left(\frac{n-1}{2}+i\right)\Big/\Gamma\left(\frac{p-1}{2}+i\right)\cdot\Gamma\left(\frac{n-p}{2}\right).$$

作变换

$$t=\frac{n-p}{p-1}\,\frac{w}{1-w}$$

则

$$J(t\to w)=\frac{n-p}{p-1}(1-w)^{-2}$$

于是当 λ 给定后 $w=r^2$ 的条件密度为

$$f(w,\lambda)=e^{-\frac{\lambda}{2}}(1-w)^{\frac{n-p-2}{2}}\sum_{i=0}^{\infty}c_i\cdot\frac{1}{i!}\left(\frac{\lambda w}{2}\right)^{i}w^{\frac{p-1}{2}-1} \tag{4}$$

只要在(4)中关于 λ 积分,就可得到 $w=r^2$ 的无条件分布,即

$$f(w)=\int_0^{\infty}f(w,\lambda)g(\lambda)d\lambda.$$

由

$$\int_0^{\infty}\left(\frac{\lambda w}{2}\right)^{i}\frac{1}{\Gamma\left(\frac{n-1}{2}\right)}\cdot\left(\frac{\nu}{2}\right)^{\frac{n-1}{2}}\lambda^{\frac{n-3}{2}}\exp\left(-\frac{1+\nu}{2}\lambda\right)d\lambda$$

$$=\frac{\Gamma\left(\frac{n-1}{2}+i\right)}{\Gamma\left(\frac{n-1}{2}\right)}(1-\bar\rho^2)^{\frac{n-1}{2}}(\bar\rho^2 w)^{i}$$

这里利用了 $1+\nu=\bar\rho^{-2}$, $\frac{\nu}{1+\nu}=1-\bar\rho^2$.

所以 $w=r^2$ 的无条件分布密度为

$$(1-\bar\rho^2)^{\frac{n-1}{2}}(1-w)^{\frac{n-p-2}{2}}\sum_{i=0}^{\infty}\frac{(\bar\rho^2 w)^{i}}{i!}c_i$$

$$\cdot \frac{\Gamma\left(\frac{n-1}{2}+i\right)}{\Gamma\left(\frac{n-1}{2}\right)} w^{\frac{p-1}{2}-1} \tag{5}$$

最后再令 $w=\bar{r}^2$，$J(\omega \to \bar{r})=2\bar{r}$，再将 c_i 表达式代入即可证得本定理。∎

第五章　统计假设检验

统计假设检验是统计推断的基本问题之一．在一元统计中，对于正态总体 $N(\mu,\sigma^2)$ 的参数 μ 和 σ^2 进行了各种检验，作为一元的推广．本章主要介绍多元正态总体 $N_p(\boldsymbol{\mu},\Sigma)$ 的假设检验问题．§5.1～§5.3 是介绍正态总体均值向量的各种检验，§5.4 是介绍多项分布的参数的大样本推断及非正态总体均值向量的检验，§5.5 是介绍正态总体协方差矩阵的检验．

§5.1　协方差阵已知时均值向量的检验

（一）似然比统计量

在数理统计中有关分布的参数的假设检验，通常是利用极大似然原理导出似然比统计量进行检验，在多元统计分析中几乎所有重要的假设检验都是利用极大似然原理给出的，因此在讨论具体的检验问题之前，首先对极大似然原理作一个回顾．

设 p 维总体的密度函数为 $f(\boldsymbol{x},\boldsymbol{\theta})$，其中 $\boldsymbol{\theta}$ 是未知参数，它取值于参数空间 Θ，又设 Θ_0 是 Θ 的一个子集．我们希望对下列假设

$$H_0:\boldsymbol{\theta}\in\Theta_0,\quad H_1:\boldsymbol{\theta}\notin\Theta_0 \tag{1.1}$$

作出判断，这就是假设检验问题．我们称 H_0 为原假设，H_1 为备择假设．

从该总体独立抽取了容量为 n 的样本 $\boldsymbol{x}_i(i=1,2,\cdots,n)$，记它的似然函数为

$$L(\boldsymbol{\theta})=L(\boldsymbol{x}_1,\cdots\boldsymbol{x}_n;\boldsymbol{\theta})=\prod_{i=1}^{n}f(\boldsymbol{x}_i,\boldsymbol{\theta})$$

引进统计量

$$\lambda = \frac{\max\limits_{\theta \in \Theta_0} L(\theta)}{\max\limits_{\theta \in \Theta} L(\theta)} \tag{1.2}$$

显然 λ 是不依赖参数 θ,而仅是样本 $x_1, \cdots, x_n$ 的函数,从而它是一个统计量,习惯上称 λ 为似然比统计量,由于 $\Theta_0 \subset \Theta$, 从而 $0 \leqslant \lambda \leqslant 1$,

由极大似然原理知,如果 λ 取值较小, 这说明当 H_0 为真时观察到的样本 $x_1, x_2, \cdots, x_n$ 的概率比 H_0 不真时观察到的样本 $x_1, x_2, \cdots, x_n$ 的概率要小得多,此时,我们有理由怀疑假设 H_0 不真. 所以从似然比出发,(1.1)的检验问题是当下式成立时拒绝 H_0,

$$\lambda(x_1, x_2, \cdots, x_n) \leqslant \lambda_0$$

其中 λ_0 是临界值,它是由给定的显著水平 α,按 $P_\theta(\lambda \leqslant \lambda_0) \leqslant \alpha$, 对一切 $\theta \in \Theta_0$ 来确定的. 由此可见,要想确定 λ_0, 就必须求出似然比统计量 λ 在假设 H_0 成立时的分布.

(二) 单总体的情况

设 $x_1, \cdots, x_n$ 是取自 p 元正态总体 $N_p(\mu, \Sigma)$ 容量为 n 的样本,其中 μ 为均值向量, Σ 为正定的协方差矩阵,且为已知的,我们的问题是在 Σ 已知的情况下,通过样本 $x_1, \cdots, x_n$, 来检验总体的均值向量 μ 是否等于事先指定的 μ_0. 即检验假设

$$H_0: \mu = \mu_0, \qquad H_1: \mu \neq \mu_0 \tag{1.3}$$

显然, 此时 $\Theta = \{\mu: \mu$ 为任意的 p 维实向量$\}$ $\Theta_0 = \{\mu: \mu = \mu_0\}$ 似然函数为

$$\begin{aligned} L(\mu) &= L(x_1, \cdots, x_n; \mu) \\ &= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)'\Sigma^{-1}(x_i - \mu)\right\} \\ &= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}[\Sigma^{-1}(A + n(\bar{x} - \mu)(\bar{x} - \mu)')]\right\} \end{aligned}$$

其中

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$A = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})'$$

由于Σ已知,$L(\mu)$当且仅当在 $\mu = \bar{x}$ 处达到最大值,即

$$\max_{\mu \in \Theta} L(\mu) = L(\bar{x})$$

$$= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}A)\right\}$$

而

$$\max_{\mu \in \Theta_0} L(\mu) = L(\mu_0)$$

$$= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}A) - \frac{1}{2}\operatorname{tr}[n\Sigma^{-1}(\bar{x} - \mu_0)(\bar{x} - \mu_0)']\right\}$$

于是似然比统计量为

$$\lambda^* = \frac{\max\limits_{\mu \in \Theta_0} L(\mu)}{\max\limits_{\theta \in \Theta} L(\mu)} = \exp\left\{-\frac{1}{2}\operatorname{tr}[n\Sigma^{-1}(\bar{x} - \mu_0)(\bar{x} - \mu_0)']\right\}$$

$$= \exp\left\{-\frac{1}{2} n(\bar{x} - \mu_0)'\Sigma^{-1}(\bar{x} - \mu_0)\right\} = \exp\left\{-\frac{1}{2}\lambda\right\}$$

其中

$$\lambda = n(\bar{x} - \mu_0)'\Sigma^{-1}(\bar{x} - \mu_0) \tag{1.4}$$

显然 λ^* 是 λ 的严格单调下降函数，因此对于给出的显著水平 α，当 $\lambda^* \leqslant \lambda^*(\alpha)$ 就等价于 $\lambda \geqslant \lambda(\alpha)$,于是上述的似然比统计量等价于$\lambda$ 统计量. 由第三章定理 1.1 知在 H_0 为真时 $\lambda = n(\bar{x} - \mu_0)'\Sigma^{-1}(\bar{x} - \mu_0) \sim \chi^2(p)$. 从而关于 (1.3) 的假设检验有如下定理:

定理 1.1 设 $x_1, \cdots, x_n$ 是取自正态总体 $N_p(\mu, \Sigma)$ 的样本($\Sigma > 0$ 且已知, $n > p$),对于显著水平 α, 则拒绝 (1.3) 中原假设

H_0 的拒绝域是

$$\lambda = n(\bar{x} - \mu_0)'\Sigma^{-1}(\bar{x} - \mu_0) \geqslant \chi^2_{1-\alpha}(p) \tag{1.5}$$

其中 $\chi^2_{1-\alpha}(p)$为自由度是 p 的 χ^2 变量的$1-\alpha$ 分位数，即$P\{\chi^2(p) \leqslant \chi^2_{1-\alpha}(p)\} = 1 - \alpha$，习惯上称 $\chi^2_{1-\alpha}(p)$为临界值。

检验的具体步骤可归结如下：

(i) 根据样本值计算样本的均值向量 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$，计算已知协方差阵 Σ 的逆矩阵 Σ^{-1}，然后代入统计量(1.4)得到 λ 的一个具体值；

(ii) 由给定的显著水平 α，查自由度为 p 的 χ^2 表求得临界值 $\chi^2_{1-\alpha}(p)$；

(iii) 根据(i)计算的 λ 值是否满足

$$n(\bar{x} - \mu_0)'\Sigma^{-1}(\bar{x} - \mu_0) \geqslant \chi^2_{1-\alpha}(p)$$

而决定拒绝还是接受原假设 H_0。

当拒绝 H_0 后，我们希望利用样本均值向量 $\bar{x}$ 来估计 μ 的取值范围，为此考虑不等式

$$n(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu) \leqslant \chi^2_{1-\alpha}(p) \tag{1.6}$$

由于 $n(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu) \sim \chi^2(p)$，因此 $\bar{x}$ 满足(1.6)的概率是$1-\alpha$，对于已实现的样本均值向量 $\bar{x}$ 而言(1.6)式给出了一个 μ 的集合，称这个集合为 μ 的一个置信区域，置信系数为 $1-\alpha$。此区域是一个以 $\bar{x}$ 为中心的椭球，形状依赖于 Σ，大小则取决于 $\chi^2_{1-\alpha}(p)$。

（三） 双总体的情况

设有两个 p 元正态总体 $N_p(\mu,\Sigma)$ 及 $N_p(\nu,\Sigma)$ 其中 $\Sigma > 0$ 且已知，又设 $x_1,\cdots,x_{n_1}$ 和 $y_1,\cdots,y_{n_2}$ 分别为独立取自这两个正态总体的样本，且这两个样本间也是独立抽取的，我们考虑如下的假设检验问题

$$H_0: \mu = \nu, \quad H_1: \mu \neq \nu \tag{1.7}$$

显然，与其相应的参数空间可写成

$\Theta = \{(\mu,\nu): \mu,\nu$ 为两个任意的 p 维实向量$\}$

$$\Theta_0 = \{(\boldsymbol{\mu},\boldsymbol{\nu}):\boldsymbol{\mu} = \boldsymbol{\nu}\}$$

记

$$\bar{\boldsymbol{x}} = \frac{1}{n_1}\sum_{i=1}^{n_1}\boldsymbol{x}_i,\quad \bar{\boldsymbol{y}} = \frac{1}{n_2}\sum_{i=1}^{n_2}\boldsymbol{y}_i$$

$$A_1 = \sum_{i=1}^{n_1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})',\quad A_2 = \sum_{i=1}^{n_2}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})'$$

此时似然函数为

$$\begin{aligned} L(\boldsymbol{\mu},\boldsymbol{\nu}) &= L(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{n_1};\boldsymbol{y}_1,\cdots,\boldsymbol{y}_{n_2};\boldsymbol{\mu},\boldsymbol{\nu}) \\ &= (2\pi)^{-\frac{p(n_1+n_2)}{2}}|\Sigma|^{-\frac{n_1+n_2}{2}} \\ &\quad\cdot \exp\Big\{-\frac{1}{2}\operatorname{tr}[\Sigma^{-1}(A_1 + A_2)] \\ &\quad - \frac{n_1}{2}\operatorname{tr}[\Sigma^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'] \\ &\quad - \frac{n_2}{2}\operatorname{tr}[\Sigma^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{\nu})(\bar{\boldsymbol{y}} - \boldsymbol{\nu})']\Big\} \end{aligned}$$

类似于单总体情况可以导出似然比统计量为

$$\begin{aligned} \lambda^* &= \exp\left\{-\frac{1}{2}\cdot\frac{n_1\cdot n_2}{n_1+n_2}(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}})'\Sigma^{-1}(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}})\right\} \\ &= \exp\left\{-\frac{1}{2}\lambda\right\} \end{aligned}$$

其中

$$\lambda = \frac{n_1 n_2}{n_1+n_2}(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}})'\Sigma^{-1}(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}) \tag{1.8}$$

显然，统计量 λ^* 与 λ 是等价的。由于 $\bar{\boldsymbol{x}}$ 与 $\bar{\boldsymbol{y}}$ 相互独立，所以 $(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}) \sim N_p\left(\boldsymbol{\mu} - \boldsymbol{\nu},\ \frac{n_1+n_2}{n_1\cdot n_2}\Sigma\right)$。在 H_0 为真时有 $\lambda \sim \chi^2(p)$，从而检验(1.7)就有如下定理。

定理 1.2 设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{n_1}$ 取自正态总体 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$，$\boldsymbol{y}_1,\cdots,\boldsymbol{y}_{n_2}$ 取自正态总体 $N_p(\boldsymbol{\nu},\Sigma)$ 的独立样本，其中 $\boldsymbol{\Sigma} > 0$ 且已知。对于显著水平 α，拒绝(1.7)中原假设 H_0 的拒绝域是

$$\lambda = \frac{n_1 \cdot n_2}{n_1 + n_2} (\bar{x} - \bar{y})' \Sigma^{-1} (\bar{x} - \bar{y}) \geqslant \chi^2_{1-\alpha}(p) \tag{1.9}$$

其中 $\chi^2_{1-\alpha}(p)$ 满足 $P(\chi^2(p) \leqslant \chi^2_{1-\alpha}(p)) = 1 - \alpha$.

检验步骤与(二)相同.

如果拒绝了 H_0,即认为 $\boldsymbol{\mu}$ 与 $\boldsymbol{\nu}$ 有显著差异，我们还希望通过样本均值向量来估计 $\boldsymbol{\mu} - \boldsymbol{\nu}$ 的范围,类似于(二)中的讨论,$(\boldsymbol{\mu} - \boldsymbol{\nu})$ 的 $1-\alpha$ 的置信区域可取为

$$\frac{n_1 \cdot n_2}{n_1 + n_2} [(\bar{x} - \bar{y}) - (\boldsymbol{\mu} - \boldsymbol{\nu})]' \Sigma^{-1} [(\bar{x} - \bar{y}) - (\boldsymbol{\mu} - \boldsymbol{\nu})] \leqslant \chi^2_{1-\alpha}(p) \tag{1.10}$$

此区域是以 $(\bar{x} - \bar{y})$ 为中心的一个椭球.

§5.2 协方差阵未知时均值向量的检验

上节给出了当总体协方差阵已知时总体均值向量的检验问题。但在实际问题中协方差阵通常并不知道。在一元统计中当方差 σ^2 未知时，检验均值的问题采用的是 t-检验法；即检验统计量服从 t 分布。在多元正态的情况下，利用似然比方法可以得到类似的 Hotelling-T^2 统计量

(一) 单总体的情况

设 $x_1, \cdots, x_n$ 取自 p 维正态总体 $N_p(\boldsymbol{\mu}, \Sigma)$ 的容量为 $n(n > p)$的样本,其中 $\boldsymbol{\mu}$ 和 Σ 都未知，且 $\Sigma > 0$. 我们要检验该总体的均值向量是否等于已知向量 $\boldsymbol{\mu}_0$,即检验假设

$$H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0, \quad H_1: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0 \tag{2.1}$$

显然,此时 $\Theta = \{(\boldsymbol{\mu}, \Sigma): \boldsymbol{\mu}$ 为任意的 p 维实向量, $\Sigma > 0\}$, $\Theta_0 = \{(\boldsymbol{\mu}, \Sigma): \boldsymbol{\mu} = \boldsymbol{\mu}_0, \Sigma > 0\}$. 由第四章的定理 1.1 有

$$\max_{(\boldsymbol{\mu}, \Sigma) \in \Theta} L(\boldsymbol{\mu}, \Sigma) = L\left(\bar{x}, \frac{1}{n} A\right) = (2\pi)^{-\frac{np}{2}} \left| \frac{1}{n} A \right|^{-\frac{n}{2}} \exp\left\{-\frac{np}{2}\right\}$$

其中

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\ A=\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'.$$

用类似方法可求得

$$\max_{(\mu,\Sigma)\in\Theta_0}L(\mu,\Sigma)=L\left(\mu_0,\frac{1}{n}A^*\right)$$

$$=(2\pi)^{-\frac{np}{2}}\left|\frac{1}{n}A^*\right|^{-\frac{n}{2}}\exp\left\{-\frac{np}{2}\right\}$$

其中

$$A^*=\sum_{i=1}^{n}(x_i-\mu_0)(x_i-\mu_0)'$$

$$=A+n(\bar{x}-\mu_0)(\bar{x}-\mu_0)'$$

于是似然比统计量为

$$\lambda=\frac{|A|^{\frac{n}{2}}}{|A+n(\bar{x}-\mu_0)(\bar{x}-\mu_0)'|^{\frac{n}{2}}}$$

注意到分块矩阵的行列式公式，我们取行列式

$$\begin{vmatrix} A & \sqrt{n}(\bar{x}-\mu_0) \\ -\sqrt{n}(\bar{x}-\mu_0)' & 1 \end{vmatrix}$$

$$=|A|\cdot|1+n(\bar{x}-\mu_0)'A^{-1}(\bar{x}-\mu_0)|$$

$$=|A+n(\bar{x}-\mu_0)(\bar{x}-\mu_0)'|$$

由上式中的后一个等式可得

$$\frac{|A|}{|A+n(\bar{x}-\mu_0)(\bar{x}-\mu_0)'|}$$

$$=[1+n(\bar{x}-\mu_0)'A^{-1}(\bar{x}-\mu_0)]^{-1}$$

所以有

$$\lambda=(1+n(\bar{x}-\mu_0)'A^{-1}(\bar{x}-\mu_0))^{-\frac{n}{2}}$$

$$=\left(1+\frac{T^2}{n-1}\right)^{-\frac{n}{2}}$$

其中

$$T^2 = n(n-1)(\bar{x}-\mu_0)'A^{-1}(\bar{x}-\mu_0) \tag{2.2}$$

(2.2)正是第三章中讨论的 Hotelling 的 T^2 统计量。而 λ 是 T^2 的严格单调下降函数，因此 $\lambda \leqslant \lambda(\alpha)$ 相当于 $T^2 \geqslant T^2(\alpha)$，由第三章定理 3.1 知道在 H_0 为真时有

$$\frac{n-p}{p} \cdot \frac{T^2}{n-1} \sim F(p, n-p)$$

于是有

定理 2.1 设 $x_1, \cdots, x_n$ 是取自正态总体 $N_p(\mu, \Sigma)$的样本 $(\Sigma > 0, n > p)$，其中 μ 和 Σ 都未知，对于显著水平 α，拒绝(2.1)中原假设 H_0 的拒绝域为

$$T^2 = n(n-1)(\bar{x}-\mu_0)'A^{-1}(\bar{x}-\mu_0) \geqslant T^2(\alpha) = \frac{(n-1)p}{n-p}F_{1-\alpha}(p, n-p) \tag{2.3}$$

其中 $F_{1-\alpha}(p, n-p)$ 是第一自由度为 p，第二自由度为 $n-p$ 的 F 分布的 $1-\alpha$ 分位点。

例 2.1 对 20 名健康女性的汗水进行化验，数据列在表 2.1. 其中 x_1 = 排汗量，x_2 = 汗水中钠的含量，x_3 = 汗水中钾的含量，为了探索新的诊断技术，需要检验假设 $H_0: \mu = (4, 50, 10)'$. 显著水平 $\alpha = 0.10$，并设 $(x_1, x_2, x_3)' \sim N_3(\mu, \Sigma)$，$\mu$，$\Sigma$ 都未知.

表 2.1 20 名女性的汗水化验数据

试者序号	1	2	3	4	5	6	7	8	9	10
x_1	3.7	5.7	3.8	3.2	3.1	4.6	2.4	7.2	6.7	5.4
x_2	48.5	65.1	47.2	53.2	55.5	36.1	24.8	33.1	47.4	54.1
x_3	9.3	8.0	10.9	12.0	9.7	7.9	14.0	7.6	8.5	11.3

试者序号	11	12	13	14	15	16	17	18	19	20
x_1	3.9	4.5	3.5	4.5	1.5	8.5	4.5	6.5	4.1	5.5
x_2	36.9	58.8	27.8	40.2	13.5	56.4	71.6	52.8	44.1	40.9
x_3	12.7	12.3	9.8	8.4	10.1	7.1	8.2	10.9	11.2	9.4

由表 2.1 的数据可计算得

$$\bar{x} = \begin{pmatrix} 4.640 \\ 45.400 \\ 9.965 \end{pmatrix},\ A = 19 \cdot \begin{pmatrix} 2.879 & 10.002 & -1.810 \\ 10.002 & 199.798 & -5.627 \\ -1.810 & -5.627 & 3.628 \end{pmatrix}$$

$$A^{-1} = \frac{1}{19} \cdot \begin{pmatrix} 0.586 & -0.022 & 0.258 \\ -0.022 & 0.006 & -0.002 \\ 0.258 & -0.002 & 0.402 \end{pmatrix}$$

将它们代入统计量(2.2)得 $T^2 = 9.74$，对于 $\alpha = 0.10$，查自由度 $(p, n-p) = (3, 17)$ 的 F 表得 $F_{0.90}(3,17) = 2.44$，而 $T^2(0.10) = \frac{19 \times 3}{17} F_{0.90}(3,17) = \frac{19 \times 3}{17} \times 2.44 = 8.18$。显然，$T^2 = 9.74 > 8.18 = T^2(0.10)$，于是以显著水平0.10拒绝 H_0。

如果拒绝了 H_0，我们还可以由给定的样本来估计总体均值向量 μ 的 $1-\alpha$ 置信域。对于给定的显著水平有

$$\begin{aligned} P &\left\{ n(n-1)(\bar{x}-\mu)' A^{-1}(\bar{x}-\mu) \leqslant \frac{(n-1)p}{n-p} F_{1-\alpha}(p, n-p) \right\} \\ &= P \left\{ (\bar{x}-\mu)' A^{-1}(\bar{x}-\mu) \leqslant \frac{p}{n(n-p)} F_{1-\alpha}(p, n-p) \right\} = 1-\alpha \end{aligned}$$

所以 μ 的 $1-\alpha$ 置信区域为

$$(\bar{x}-\mu)' A^{-1}(\bar{x}-\mu) \leqslant \frac{p}{n(n-p)} F_{1-\alpha}(p, n-p)$$

它是一个以 $\bar{x}$ 为中心，形状依赖于 A 的椭球。

在构造均值向量 μ 的置信区域时，在应用上更需要考虑 μ 的各个分量 μ_i 的同时置信区间 (Simultaneous Confidence Intervals) 问题。设 $x \sim N_p(\mu, \Sigma)$，考虑线性组合

$$z = l'x = l_1 x_1 + \cdots + l_p x_p$$

其中 $\boldsymbol{l}$ 为任意的非零 p 维实向量。 由第二章定理 2.1 知 $z \sim N(\boldsymbol{l}'\boldsymbol{\mu}, \boldsymbol{l}'\Sigma\boldsymbol{l})$，设 $\boldsymbol{x}_i, i=1,2,\cdots,n$ 为正态总体 $N_p(\boldsymbol{\mu},\Sigma)$的样本，相对应于 z 的样本为

$$z_i = \boldsymbol{l}'\boldsymbol{x}_i = l_1x_{1i} + \cdots + l_px_{pi} \quad i=1,2,\cdots,n.$$

且样本的均值和样本方差分别为

$$\bar{z} = \boldsymbol{l}'\bar{\boldsymbol{x}}, \quad \hat{\sigma}_z^2 = \boldsymbol{l}'S\boldsymbol{l}$$

其中 $\bar{\boldsymbol{x}}, S$ 是样本 $\boldsymbol{x}_i, i=1,2,\cdots,n$ 的样本均值向量和样本协方差阵。

对于任意的 $\boldsymbol{l}$，根据 $\boldsymbol{l}'\boldsymbol{\mu}$ 的置信区间能够得到所要的同时置信区间。事实上，对于任意固定的 $\boldsymbol{l}$ 和未知的 $\sigma_z^2 = \boldsymbol{l}'\Sigma\boldsymbol{l}$，由一元统计知道基于 t-统计量的 $\mu_z = \boldsymbol{l}'\boldsymbol{\mu}$ 的 $1-\alpha$ 置信区间为

$$\boldsymbol{l}'\bar{\boldsymbol{x}} - t_{1-\frac{\alpha}{2}}(n-1)\frac{\sqrt{\boldsymbol{l}'S\boldsymbol{l}}}{\sqrt{n}} \leqslant \boldsymbol{l}'\boldsymbol{\mu}$$

$$\leqslant \boldsymbol{l}'\bar{\boldsymbol{x}} + t_{1-\frac{\alpha}{2}}(n-1)\frac{\sqrt{\boldsymbol{l}'S\boldsymbol{l}}}{\sqrt{n}}$$

若取 $\boldsymbol{l}' = (0,\cdots,0,1,0,\cdots,0)$，即 $\boldsymbol{l}$ 的第 i 个分量为 1 其余分量皆为 0，则 $\boldsymbol{l}'\boldsymbol{\mu} = \mu_i$，上式即为第 i 个分量 μ_i 的 $1-\alpha$ 置信区间。显然，选择不同的 $\boldsymbol{l}$ 就可以得到 $\boldsymbol{\mu}$ 的不同线性组合的 $1-\alpha$ 置信区间。但需注意，对不同的 $\boldsymbol{l}$，$\boldsymbol{l}'\boldsymbol{\mu}$ 同时落入各自的 $1-\alpha$ 置信区间的总置信度就不再是 $1-\alpha$，而要比 $1-\alpha$ 低。为了保证总置信度不小于$1-\alpha$，下面介绍 Scheffe 构造同时置信区间的方法。

对于给定的样本 $\boldsymbol{x}_i, i=1,2,\cdots,n$ 和固定的 $\boldsymbol{l}$，置信区间

$$\left[\boldsymbol{l}'\bar{\boldsymbol{x}} - t^{1-\frac{\alpha}{2}}(n-1)\cdot\frac{\sqrt{\boldsymbol{l}'S\boldsymbol{l}}}{\sqrt{n}}, \boldsymbol{l}'\bar{\boldsymbol{x}} + t_{1-\frac{\alpha}{2}}(n-1)\cdot\frac{\sqrt{\boldsymbol{l}'S\boldsymbol{l}}}{\sqrt{n}}\right]$$

是使

$$t^2 = \frac{n[\boldsymbol{l}'(\bar{\boldsymbol{x}}-\boldsymbol{\mu})]^2}{\boldsymbol{l}'S\boldsymbol{l}} \leqslant t^2_{1-\frac{\alpha}{2}}(n-1)$$

成立的全体 $\boldsymbol{l}'\boldsymbol{\mu}$ 值的范围。为求 $\boldsymbol{l}'\boldsymbol{\mu}$ 的同时置信区间，我们设法

寻找一个与 $\boldsymbol{l}$ 无关的统计量 $T(\boldsymbol{x})$,使对于所有 $\boldsymbol{l}$ 均有

$$\frac{n[\boldsymbol{l}'(\bar{\boldsymbol{x}}-\boldsymbol{\mu})]^2}{\boldsymbol{l}'S\boldsymbol{l}} \leqslant T(\boldsymbol{x})$$

然后由 $T(\boldsymbol{x})$ 的置信区间来求得 $\boldsymbol{l}'\boldsymbol{\mu}$ 的同时置信区间。这就导致了求 t^2 的最大值:

$$\max_{l} t^2 = \max_{l} \frac{n[\boldsymbol{l}'(\bar{\boldsymbol{x}}-\boldsymbol{\mu})]^2}{\boldsymbol{l}'S\boldsymbol{l}}$$

由第一章定理 3.13 有

$$\max_{l} \frac{n[\boldsymbol{l}'(\bar{\boldsymbol{x}}-\boldsymbol{\mu})]^2}{\boldsymbol{l}'S\boldsymbol{l}} = n(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'S^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}) = T^2$$

注意上式中 S 为样本协方差阵即

$$(n-1)S = \sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})' = A$$

对于给定的 α 当

$$T^2 = n(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'S^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}) \leqslant \frac{(n-1)p}{n-p}F_{1-\alpha}(p,n-p)$$

时就意味着对一切 $\boldsymbol{l}$ 有

$$\frac{n[\boldsymbol{l}'\bar{\boldsymbol{x}}-\boldsymbol{l}'\boldsymbol{\mu}]^2}{\boldsymbol{l}'S\boldsymbol{l}} \leqslant \frac{(n-1)p}{n-p}F_{1-\alpha}(p,n-p)$$

即对一切 $\boldsymbol{l}$ 有

$$\boldsymbol{l}'\bar{\boldsymbol{x}} - \sqrt{\frac{(n-1)p}{n-p}F_{1-\alpha}(p,n-p)}\cdot\sqrt{\frac{\boldsymbol{l}'S\boldsymbol{l}}{n}} \leqslant \boldsymbol{l}'\boldsymbol{\mu}$$

$$\leqslant \boldsymbol{l}'\bar{\boldsymbol{x}} + \sqrt{\frac{(n-1)p}{n-p}F_{1-\alpha}(p,n-p)}\cdot\sqrt{\frac{\boldsymbol{l}'S\boldsymbol{l}}{n}}$$

或者说对所有 $\boldsymbol{l}$,区间

$$\left[\boldsymbol{l}'\bar{\boldsymbol{x}} - \sqrt{\frac{p(n-1)}{n(n-p)}F_{1-\alpha}(p,n-p)\cdot\boldsymbol{l}'S\boldsymbol{l}}\right.,$$

$$\left.\boldsymbol{l}'\bar{\boldsymbol{x}} + \sqrt{\frac{p(n-1)}{n(n-p)}F_{1-\alpha}(p,n-p)\boldsymbol{l}'S\boldsymbol{l}}\right] \qquad (2.4)$$

包含$l'\boldsymbol{\mu}$的概率为1-α。为了方便，我们将区间(2.4)称为T^2-区间。若依序取$l'=(1,0,\cdots,0)$;$l'=(0,1,0,\cdots,0)$;$\cdots$;$l'=(0,0,\cdots,0,1)$，就可得到$\boldsymbol{\mu}$的所有p个分量μ_i的置信度为1-α的T^2-区间

$$\left[\bar{x}_i-\sqrt{\frac{p(n-1)}{n(n-p)}F_{1-\alpha}(p,n-p)\cdot s_{ii}},\right.$$

$$\left.\bar{x}_i+\sqrt{\frac{p(n-1)}{n(n-p)}\cdot F_{1-\alpha}(p,n-p)\cdot s_{ii}}\right] \tag{2.5}$$

其中$\bar{x}_i$为$\bar{\boldsymbol{x}}$的第i个分量，s_{ii}为S的(i,i)元素。类似地取$l'=(0,\cdots,0,l_i,0,\cdots,0,l_k,0,\cdots,0)$，其中$l_i=1,l_k=-1,(i<k,k=2,\cdots,n)$。此时$l'Sl=s_{ii}-2s_{ik}+s_{kk}$，我们就得到所有$(\mu_i-\mu_k)$的置信度为1-$\alpha$的$T^2$-区间

$$\left[(\bar{x}_i-\bar{x}_k)\pm\sqrt{\frac{p(n-1)}{n(n-p)}F_{1-\alpha}(p,n-p)(s_{ii}-2s_{ik}+s_{kk})}\right] \tag{2.6}$$

在实际应用上我们常常对为数不多的同时置信区间感兴趣。在这种情况下，我们可能得到比T^2-区间更好的结果，事实上，我们考虑m个线性组合$l_i'\boldsymbol{\mu},i=1,2,\cdots,m$，且$m$较小。现在要构造这$m$个线性组合的同时置信区间，利用Bonferroni概率不等式，设$B_i,i=1,2,\cdots,m$为m个随机事件，且$P(B_i)=1-\alpha_i$则有

$$P\left(\bigcap_{i=1}^{m}B_i\right)=1-P\left(\overline{\bigcap_{i=1}^{m}B_i}\right)=1-P\left(\bigcup_{i=1}^{m}\bar{B}_i\right)$$

$$\geqslant 1-\sum_{i=1}^{m}P(\bar{B}_i)=1-\sum_{i=1}^{m}\alpha_i$$

对于每个t-区间

$$\left[\bar{x}_i\pm t_{1-\frac{\alpha_i}{2}}(n-1)\cdot\sqrt{\frac{s_{ii}}{n}}\right],\quad i=1,2,\cdots,m$$

其中$\alpha_i=\dfrac{\alpha}{m}$，令

$$B_i = \left\{\mu_i \in \left[\bar{x}_i \pm t_{1-\frac{\alpha}{2m}}(n-1)\cdot\sqrt{\frac{s_{ii}}{n}}\right]\right\}$$

则有 $P(B_i) = 1 - \frac{\alpha}{m}$. 由 Bonferroni 不等式得到 $P\left(\bigcap_{i=1}^{m} B_i\right) \geqslant 1-\alpha$. 特别取 $m = p$ 就可得到均值向量的 p 个分量 μ_i, $i = 1, 2, \cdots, p$. 的 $1-\alpha$ 同时置信区间

$$\left[\bar{x}_i \pm t_{1-\frac{\alpha}{2p}}(n-1)\cdot\sqrt{\frac{s_{ii}}{n}}\right],\quad i = 1,2,\cdots p \tag{2.7}$$

例 2.2 在例 2.1 中原假设 H_0 在 $\alpha = 0.10$ 下被拒绝，此时我们利用(2.7)式容易求得 μ_1, μ_2, μ_3 的 0.95 Bonferroni 同时置信区间。事实上，若取 $\alpha_i = \frac{0.05}{3}$ $i = 1,2,3$，因为 $n = 20$，所以查 t-分布表得

$$t_{1-\frac{0.05}{6}}(19) = t_{0.9917}(19) = 2.625$$

再利用例 2.1 的计算结果便可得到 0.95 同时置信区间为

$3.64 \leqslant \mu_1 \leqslant 5.64, 37.10 \leqslant \mu_2 \leqslant 53.70, 8.85 \leqslant \mu_3 \leqslant 11.08$

对于一般形式的 m 个线性组合 $l_i'\mu$, $i = 1, 2, \cdots, m$ 的 Bonferroni 同时置信区间为

$$\left[l_i'\bar{x} \pm t_{1-\frac{\alpha}{2m}}(n-1)\cdot\sqrt{\frac{l_i'Sl_i}{n}}\right]\quad i = 1,2,\cdots,m \tag{2.8}$$

它与 T^2-区间(2.4)具有相同形式，即都是以 $l_i'\bar{x}$ 为中心的一个区间，它们之间的长度比为

$$\frac{\text{Bonferroni 区间长}}{T^2\text{-区间长}} = \frac{t_{1-\frac{\alpha}{2m}}(n-1)}{\sqrt{\frac{p(n-1)}{n-p}F_{1-\alpha}(p,n-p)}}$$

它不依赖于统计量 $\bar{x}$ 和 S，对为数不多的 m 个 $l_i'\mu$，Bonferroni 区间总要短些。表 2.2 对某些 n 及 m 值列出了两者的长度比值。

（二） 具有等协方差阵双总体情况

设 $x_1, \cdots, x_{n_1}$ 和 $y_1, \cdots, y_{n_2}$ 分别是取自总体 $N_p(\mu, \Sigma)$ 和

表 2.2 Bonferroni 区间与 T^2-区间长度比

$$\left(1-\alpha=0.95, \alpha_i=\frac{0.05}{m}\right)$$

n \ 比值 \ $m=p$	2	4	10
15	0.88	0.69	0.29
25	0.90	0.75	0.48
50	0.91	0.78	0.58
100	0.91	0.80	0.62
∞	0.91	0.81	0.66

总体 $N_p(\nu,\Sigma)$ 的样本 $(n_1>p, n_2>p)$ 其中 $\Sigma>0$ 且未知. 两个样本相互独立. 考虑如下的检验假设问题

$$H_0: \mu=\nu, \quad H_1: \mu\neq\nu \tag{2.9}$$

类似于前面,用似然比方法可导出检验统计量

$$T^2=\frac{n_1n_2}{n_1+n_2}(n_1+n_2-2)(\bar{x}-\bar{y})'A^{-1}(\bar{x}-\bar{y}) \tag{2.10}$$

其中

$$\bar{x}=\frac{1}{n_1}\sum_{i=1}^{n_1}x_i \quad \bar{y}=\frac{1}{n_2}\sum_{i=1}^{n_2}y_i$$

$$A=A_1+A_2=\sum_{i=1}^{n_1}(x_i-\bar{x})(x_i-\bar{x})'$$

$$+\sum_{i=1}^{n_2}(y_i-\bar{y})(y_i-\bar{y})'$$

当 H_0 为真时, $T^2\sim T^2(p, n_1+n_2-2)$,由第三章定理 3.1 有

$$\frac{n_1+n_2-p-1}{p(n_1+n_2-2)}\cdot T^2\sim F(p, n_1+n_2-p-1)$$

于是有如下定理.

定理 2.2 设 $x_1,\cdots,x_{n_1}$ 取自正态总体 $N_p(\mu,\Sigma)$ 的样本. $y_1,\cdots,y_{n_2}$ 取自正态总体 $N_p(\nu,\Sigma)$ 的样本 ($\Sigma>0$ 且未知;

$n_1 > p; n_2 > p$)。两样本相互独立。对于显著水平 α 则拒绝(2.9)中原假设 H_0 的拒绝域为

$$\frac{(n_1+n_2-p-1)n_1n_2}{p(n_1+n_2)}(\bar{x}-\bar{y})'A^{-1}(\bar{x}-\bar{y})$$
$$\geqslant F_{1-\alpha}(p, n_1+n_2-p-1) \tag{2.11}$$

当拒绝了 H_0,相应地 $\mu-\nu$ 的 1-α 置信区域是

$$\frac{(n_1+n_2-p-1)n_1\cdot n_2}{p(n_1+n_2)}[(\bar{x}-\bar{y})-(\mu-\nu)]'A^{-1}[(\bar{x}-\bar{y})-(\mu-\nu)]$$
$$\leqslant F_{1-\alpha}(p, n_1+n_2-p-1)$$

它是以 $(\bar{x}-\bar{y})$ 为中心的一个椭球。

例 2.3 某小麦品种在甲、乙两地进行试验, 并分别抽取了 27 个植株观察: x_1 = 收割时植株的高度, x_2 = 有效分蘖数目, x_3 = 每植株平均穗长, x_4 = 每 10 穗中结实的小穗状花序的数目, x_5 = 每 10 穗的粒数, x_6 = 每 10 穗的粒重。 希望考察两地的平均产量是否相同,即 $\mu=\nu$ 是否成立。根据所观测到的数据分别计算得样本平均向量及样本协方差阵:

$$\bar{x}' = (84.89, 186.30, 9.74, 13.46, 304.37, 13.63)$$
$$\bar{y}' = (77.14, 167.19, 10.46, 13.10, 361.56, 14.76)$$
$$A = A_1 + A_2$$
$$= 52\cdot\begin{pmatrix} 1143.07 & & & & & \\ 57.40 & 3.84 & & & & \\ 70.16 & 4.25 & 25.54 & & & \\ 79.48 & 0.66 & 23.62 & 326.56 & & \\ 15.28 & 0.77 & 1.18 & 2.40 & 0.30 & \\ 21.60 & 1.04 & 2.56 & 4.14 & 0.39 & 0.83 \end{pmatrix}$$

设两个六元总体为正态总体且有相同的正定协方差阵。 根据(2.11)式计算得

$$\frac{(n_1+n_2-p-1)n_1n_2}{p(n_1+n_2)}(\bar{x}-\bar{y})'A^{-1}(\bar{x}-\bar{y})$$

$$=\frac{47\times 27\times 27}{6\times 54}\cdot 0.3099$$

$$=32.77$$

对于 $\alpha=0.01$ 查自由度为 $(p,n_1+n_2-p-1)=(6,47)$ 的 F 分布表得 $F_{0.99}(6,47)=3.21$,显然 $32.77>3.21$,所以两正态总体的均值向量有显著差异.

类似于单总体的情况,当拒绝了 H_0 后在实用上我们更感兴趣的是两总体均值向量差 $\boldsymbol{\mu}-\boldsymbol{\nu}$ 的分量的同时置信区间.

设 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\Sigma),\boldsymbol{y}\sim N_p(\boldsymbol{\nu},\Sigma)(\Sigma>0$ 未知),考虑线性组合 $v=\boldsymbol{l}'\boldsymbol{x}$ 和 $w=\boldsymbol{l}'\boldsymbol{y}$,$\boldsymbol{l}$ 为任意非零 p 维实向量. 设 $\boldsymbol{x}_i(i=1,2,\cdots,n_1)$ 和 $\boldsymbol{y}_i(i=1,2,\cdots,n_2)$ 分别为两总体的样本,它们的样本均值向量及样本协方差阵分别记为

$$\bar{\boldsymbol{x}}=\frac{1}{n_1}\sum_{i=1}^{n_1}\boldsymbol{x}_i\quad \bar{\boldsymbol{y}}=\frac{1}{n_2}\sum_{i=1}^{n_2}\boldsymbol{y}_i$$

$$S_1=\frac{A_1}{n_1-1}=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(x_i-\bar{\boldsymbol{x}})'$$

$$S_2=\frac{A_2}{n_2-1}=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(\boldsymbol{y}_i-\bar{\boldsymbol{y}})(\boldsymbol{y}_i-\bar{\boldsymbol{y}})'$$

相应于 v 和 w 的样本均值和样本方差为 $\bar{v}=\boldsymbol{l}'\bar{\boldsymbol{x}},\bar{w}=\boldsymbol{l}'\bar{\boldsymbol{y}}$和$\hat{\sigma}_1^2=\boldsymbol{l}'S_1\boldsymbol{l},\hat{\sigma}_2^2=\boldsymbol{l}'S_2\boldsymbol{l}$. 令

$$S=\frac{A_1+A_2}{n_1+n_2-2}=\frac{A}{n_1+n_2-2}\quad (A=A_1+A_2)$$

那么 $\boldsymbol{l}'\Sigma\boldsymbol{l}$ 的估计为 $\boldsymbol{l}'S\boldsymbol{l}$. 对于给定的 α,由一元统计知道,置信区间

$$\left[\boldsymbol{l}'(\bar{\boldsymbol{x}}-\bar{\boldsymbol{y}})\pm t_{1-\frac{\alpha}{2}}(n_1+n_2-2)\cdot\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\boldsymbol{l}'S\boldsymbol{l}}\right]$$

是使

$$t^2=\frac{[l'(\bar{x}-\bar{y})-l'(\mu-\nu)]^2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)l'Sl}\leqslant t^2_{1-\frac{\alpha}{2}}(n_1+n_2-2)$$

成立的全体 $l'(\mu-\nu)$ 值. 让 l 变化,完全相类似于单总体的情况有

$$\begin{aligned}\max_l &\frac{[l'(\bar{x}-\bar{y})-l'(\mu-\nu)]^2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)l'Sl}\\ &=[(\bar{x}-\bar{y})-(\mu-\nu)]'\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S\right]^{-1}\\ &\quad\cdot[(\bar{x}-\bar{y})-(\mu-\nu)]\\ &=\frac{n_1n_2}{n_1+n_2}\cdot(n_1+n_2-2)[(\bar{x}-\bar{y})-(\mu-\nu)]'\\ &\quad\cdot A^{-1}[(\bar{x}-\bar{y})-(\mu-\nu)]\\ &=T^2\end{aligned}$$

对于给定的 α 及一切非零 l 便可得到 $l'(\mu-\nu)$ 的 $1-\alpha$ 同时置信区间(即 T^2-区间)为

$$\left[l'(\bar{x}-\bar{y})\pm c\sqrt{l'\left(\frac{1}{n_1}+\frac{1}{n_2}\right)Sl}\right]\tag{2.12}$$

其中

$$c=\left[\frac{(n_1+n_2-2)p}{(n_1+n_2-p-1)}F_{1-\alpha}(p,n_1+n_2-p-1)\right]^{\frac{1}{2}}$$

特别取 $l'=(0,\cdots,0,1,0,\cdots,0)$, 即 l' 的第 i 个分量为 1 其余全为 0,则 $(\mu_i-\nu_i)(i=1,2,\cdots,p)$ 的 T^2-区间为

$$\left[(\bar{x}_i-\bar{y}_i)\pm c\cdot\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)s_{ii}}\right]\tag{2.13}$$

其中 $\bar{x}_i,\bar{y}_i$ 分别为 $\bar{x}$ 和 $\bar{y}$ 的第 i 个分量, c 同(2.12)中的 c, s_{ii} 为 $S=(A_1+A_2)/n_1+n_2-2$ 的 (ii) 元素. 读者不难自己去推导出 $(\mu_i-\nu_i)i=1,2,\cdots,p$ 的 Bonferroni 同时置信区间为

$$\left[(\bar{x}_i - \bar{y}_i) \pm t_{1-\frac{\alpha}{2p}}(n_1 + n_2 - 2) \cdot \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)s_{ii}}\right] \quad (2.14)$$

例 2.4　在例 2.3 中原假设被拒绝．对 $\alpha = 0.05$ 利用 (2.14) 可求出 $(\mu_i - \nu_i)$ $i = 1,2,\cdots,6$ 的 0.95 Bonferroni 同时置信区间。这里 $p = 6, n_1 = n_2 = 27$，由 t-分布表近似地得到

$$t_{1-\frac{0.05}{12}}(52) = t_{0.996}(52) \approx 2.60$$

再根据例 2.3 中 $\bar{x}_i, \bar{y}_i$ 及 $s_{ii}, i = 1,2,\cdots,6$ 的数据计算得到

$-16.18 \leqslant \mu_1 - \nu_1 \leqslant 31.68$,　　$17.72 \leqslant \mu_2 - \nu_2 \leqslant 20.50$

$-4.30 \leqslant \mu_3 - \nu_3 \leqslant 3.30$,　　$-12.43 \leqslant \mu_4 - \nu_4 \leqslant 13.15$

$-57.58 \leqslant \mu_5 - \nu_5 \leqslant -56.80$,　$-1.77 \leqslant \mu_6 - \nu_6 \leqslant -0.49$

由上面的置信区间可看出，两总体均值向量的第一、三、四分量之间的差的置信区间其上限为正，下限为负，即它们的置信区间包含零。因此对于这三个分量不能断定它们不相等的，可分别对它们用一元 t 检验方法帮助推断，但对于例 2.3，我们认为 $\boldsymbol{\mu} \neq \boldsymbol{\nu}$ 主要是第二、五、六分量有显著差异。

*(三)　具有不相等协方差阵双总体情况

设总体 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \Sigma_1)$，总体 $\boldsymbol{y} \sim N_p(\boldsymbol{\nu}, \Sigma_2)$，$\Sigma_1 > 0, \Sigma_2 > 0, \Sigma_1 \neq \Sigma_2$ 且都未知．设 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n_1}$ 和 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{n_2}$．分别为这两个总体的样本，且两样本是独立的，检验的假设问题与 (2.9) 相同。

首先考虑 $n_1 = n_2 = n$ 的情况。由于样本容量相等，经随机化处理就可用成对比较的方法。令 $\boldsymbol{z}_i = \boldsymbol{x}_i - \boldsymbol{y}_i$ $i = 1,2,\cdots,n$．则 $\boldsymbol{z}_i$ 可视为来自 p 元正态总体 $N_p(\boldsymbol{\mu} - \boldsymbol{\nu}, \Sigma_1 + \Sigma_2)$ 的样本。上述的检验问题就化为检验假设 $H_0: \boldsymbol{\mu} - \boldsymbol{\nu} = 0$，$H_1: \boldsymbol{\mu} - \boldsymbol{\nu} \neq 0$．这正是本节(一)所讨论的情况．检验的统计量为

$$T^2 = n(n-1)\bar{\boldsymbol{z}}' A^{-1} \bar{\boldsymbol{z}} \quad (2.15)$$

其中

$$\bar{\boldsymbol{z}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{y}_i)$$

$$A=\sum_{i=1}^{n}(z_i-\bar{z})(z_i-\bar{z})'$$

其次考虑 $n_1\neq n_2$ 的情况，不妨设 $n_1<n_2$，令

$$z_i=x_i-\sqrt{\frac{n_1}{n_2}}y_i+\frac{1}{\sqrt{n_1n_2}}\sum_{k=1}^{n_1}y_k-\frac{1}{n_2}\sum_{k=1}^{n_2}y_k,$$

$$i=1,2,\cdots,n_1$$

于是

$$Ez_i=\mu-\sqrt{\frac{n_1}{n_2}}\nu+\frac{n_1}{\sqrt{n_1n_2}}\nu-\frac{n_2}{n_2}\nu=\mu-\nu$$

$$\begin{aligned}\operatorname{cov}(z_i,z_j)&=E(z_i-Ez_i)(z_j-Ez_j)\\&=E\left[(x_i-\mu)-\sqrt{\frac{n_1}{n_2}}(y_i-\nu)+\frac{1}{\sqrt{n_1n_2}}\right.\\&\quad\left.\cdot\sum_{k=1}^{n_1}(y_k-\nu)-\frac{1}{n_2}\sum_{k=1}^{n_2}(y_k-\nu)\right]\\&\quad\cdot\left[(x_j-\mu)-\sqrt{\frac{n_1}{n_2}}(y_j-\nu)+\frac{1}{\sqrt{n_1n_2}}\right.\\&\quad\left.\cdot\sum_{k=1}^{n_1}(y_k-\nu)-\frac{1}{n_2}\sum_{k=1}^{n_2}(y_k-\nu)\right]'\\&=\delta_{ij}\Sigma_1+\delta_{ij}\frac{n_1}{n_2}\Sigma_2+\Sigma_2\left(-\frac{2}{n_2}+\frac{2}{n_2}\right.\\&\quad\left.\cdot\sqrt{\frac{n_1}{n_2}}+\frac{n_1}{n_1n_2}-\frac{2n_1}{\sqrt{n_1n_2}\cdot n_2}+\frac{n_2}{n_2^2}\right)\\&=\delta_{ij}\left(\Sigma_1+\frac{n_1}{n_2}\Sigma_2\right)\end{aligned}$$

其中

$$\delta_{ij}=\begin{cases}1 & i=j\\0 & i\neq j\end{cases}$$

由此可知 $z_i, i = 1,2,\cdots,n_1$ 相互独立，且 $z_i \sim N_p\left(\mu - \nu, \Sigma_1 + \frac{n_1}{n_2}\Sigma_2\right)$. 这时的检验问题仍转化为(一)所讨论的情况. 检验统计量为

$$T^2 = n_1(n_1 - 1)\bar{z}'A^{-1}\bar{z} \tag{2.16}$$

其中

$$\bar{z} = \frac{1}{n_1}\sum_{i=1}^{n_1} z_i \qquad A = \sum_{i=1}^{n_1}(z_i - \bar{z})(z_i - \bar{z})'$$

本节中所涉及的检验统计量都是 Hotelling 的 T^2 统计量.这个统计量具有一些优良的性质，首先是在非奇异变换之下的不变性. 例如,设总体 $x \sim N_p(\mu, \Sigma)$, $x_1, \cdots, x_n$ 为其样本,检验假设 $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$. 检验统计量为 $T^2 = n(n-1)(\bar{x} - \mu_0)' A^{-1}(\bar{x} - \mu_0)$. 考虑变换 $y = Cx + d$,其中 C 为 $p \times p$ 阶非奇异阵, d 为 p 维实向量. 由多元正态分布的性质知 $y \sim N_p(C\mu + d, C\Sigma C')$,相对应于 x_i 的样本为 $y_i = Cx_i + d, i = 1,2,\cdots,n$. 对总体 x 的上述检验假设等价于对总体 y 作检验假设 $H_0: \mu_y = \mu_y^{(0)}, H_1: \mu_y \neq \mu_y^{(0)}$. 这里的 $\mu_y = C\mu + d$, $\mu_y^{(0)} = C\mu_0 + d$. 记 $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $A_y = \sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})'$. 从而有检验统计量

$$T_y^2 = n(n-1)(\bar{y} - \mu_y^{(0)})' A_y^{-1}(\bar{y} - \mu_y^{(0)})$$

不难推出

$$T_y^2 = T^2 = n(n-1)(\bar{x} - \mu_0)' A^{-1}(\bar{x} - \mu_0)$$

这就是非奇异变换下的不变性,因此进行 T^2 检验时可事先对数据进行坐标平移、旋转变换，这都不影响检验结果。此外 T^2 检验还具有一致最有效性,可容许性等性质,有兴趣的读者可参看[8]的§7.2。

*§5.3 均值向量的子向量检验

前面两节是讨论正态总体的均值向量整体的各种检验问题，

本节是讨论正态总体的均值向量的子向量检验问题。

（一） 均值向量对称性检验

设总体 $x \sim N_p(\boldsymbol{\mu}, \Sigma)$，其中 $\Sigma > 0$ 未知，$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n (n > p)$ 为该总体的样本。所谓均值向量对称性的检验是指检验 $\boldsymbol{\mu}$ 的 p 个分量是否相等，即检验假设

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_p; \quad H_1: \mu_i \text{不全相等} \tag{3.1}$$

其中 $\mu_i (i = 1, 2, \cdots, p)$ 是 $\boldsymbol{\mu}$ 的第 i 个分量。

为了检验(3.1)，我们引进对照矩阵的概念。令 $\mathbf{1}_p = (1, 1, \cdots, 1)'$ 称满足 $C\mathbf{1}_p = 0$ 的 $(p-1) \times p$ 阶且秩为 $p-1$ 的矩阵 C 为对照矩阵，例如

$$C_1 = \begin{pmatrix} 1 & -1 & 0 \cdots & 0 \\ 1 & 0 & -1 \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & -1 \end{pmatrix}; \quad C_2 = \begin{pmatrix} -1 & 1 & 0 \cdots & 0 & 0 \\ 0 & -1 & 1 \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 \cdots & -1 & 1 \end{pmatrix}$$

$$C_3 = \begin{pmatrix} (1 \times 2)^{-\frac{1}{2}} & -(1 \times 2)^{-\frac{1}{2}} & 0 & \cdots\cdots & 0 \\ (2 \times 3)^{-\frac{1}{2}} & (2 \times 3)^{-\frac{1}{2}} & -2(2 \times 3)^{-\frac{1}{2}} & \cdots\cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ [(p-1)p]^{-\frac{1}{2}} & [(p-1)p]^{-\frac{1}{2}} & [(p-1) \cdot p]^{-\frac{1}{2}} & \cdots\cdots & -(p-1)[(p-1)p]^{-\frac{1}{2}} \end{pmatrix}$$

都是对照矩阵，而且 C_3 还是正交的。容易看出检验(3.1)等价于检验

$$H_0: C_1\boldsymbol{\mu} = 0, \quad H_1: C_1\boldsymbol{\mu} \neq 0 \tag{3.2}$$

或检验

$$H_0: C_2\boldsymbol{\mu} = 0, \quad H_1: C_2\boldsymbol{\mu} \neq 0 \tag{3.3}$$

注意到对于任意的 $(p-1) \times p$ 阶对照矩阵 C_1 和 C_2，必存在 $(p-1) \times (p-1)$ 阶可逆矩阵 B 使得 $C_2 = BC_1$。事实上，将 C_1 和 C_2 划分为

$$C_1 = (M_1, \boldsymbol{m}_1), \quad C_2 = (M_2, \boldsymbol{m}_2)$$

其中 M_1 和 M_2 分别为 C_1 和 C_2 前 $p-1$ 列组成的 $(p-1) \times (p-1)$ 阶矩阵，$\boldsymbol{m}_1$ 和 $\boldsymbol{m}_2$ 分别是 C_1 和 C_2 的最后一列，且不妨设

$$\mathrm{rk}(M_1) = \mathrm{rk}(M_2) = \mathrm{rk}(C_1) = \mathrm{rk}(C_2) = p - 1$$

取 $B = M_2 M_1^{-1}$，则有 $M_2 = BM_1$，又 C_1 为对照阵，所以

$$\boldsymbol{C}_1\boldsymbol{1}_p = (M_1, \boldsymbol{m}_1)\begin{pmatrix} \boldsymbol{1}_{p-1} \\ 1 \end{pmatrix} = 0$$

从而 $\boldsymbol{m}_1 = -M_1\boldsymbol{1}_{p-1}$. 类似地有 $\boldsymbol{m}_2 = -M_2\boldsymbol{1}_{p-1}$，且 $\boldsymbol{m}_2 = -M_2\boldsymbol{1}_{p-1} = -BM_1\boldsymbol{1}_{p-1} = B\boldsymbol{m}_1$，所以

$$C_2 = (M_2, \boldsymbol{m}_2) = (BM_1, B\boldsymbol{m}_1) = BC_1$$

上述的讨论说明我们考虑的对称性检验假设可通过任何一个对照矩阵 $\boldsymbol{C}$ 表示为

$$H_0: \boldsymbol{C\mu} = 0, \quad H_1: \boldsymbol{C\mu} \neq 0 \tag{3.4}$$

显然 $\boldsymbol{Cx}_i (i=1,2,\cdots,n)$ 仍独立，且每个 $\boldsymbol{Cx}_i \sim N_{p-1}(\boldsymbol{C\mu}, C\Sigma C')$. 这样，检验假设(3.4)又转化为上一节(一)中的问题. 检验统计量为

$$T^2 = n(n-1)(\boldsymbol{C\bar{x}})'(CAC')^{-1}(\boldsymbol{C\bar{x}}) \tag{3.5}$$

其中

$$\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i \quad A = \sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})' = (n-1)\cdot S$$

例 3.1 考查某地区土壤中不同深度的铅的含量是否相等. 将土壤按一定标准分为三层(表层、中层和下层)，第 i 层铅的含量为 $x_i, i = 1,2,3$，设 $\boldsymbol{x}' = (x_1, x_2, x_3) \sim N_3(\boldsymbol{\mu}, \Sigma)$. 要检验 $\boldsymbol{\mu}$ 的三个分量 μ_1, μ_2 和 μ_3 是否相等. 我们选取对照矩阵为

$$\boldsymbol{C} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix},$$

观测了 $n = 50$ 个点，即获得容量为 50 的一个样本，计算出 $\boldsymbol{Cx}$ 的样本均值和协方差阵为

$$\boldsymbol{C\bar{x}} = (13.5, 9.7)'$$

$$CSC' = \frac{1}{49}\cdot CAC' = \begin{pmatrix} 16.81 & 7.62 \\ 7.62 & 11.54 \end{pmatrix}$$

$$(CAC')^{-1} = 49\begin{pmatrix} 0.08490 & -0.05606 \\ -0.05606 & 0.12367 \end{pmatrix}$$

代入(3.5)的 T^2 统计量得 $T^2 = 621.36$. 对于 $\alpha = 0.01$,查自由度为 $(p-1, n-p+1) = (2,48)$ 的 F 表得

$$F_{0.99}(2,48) = 5.08$$

$$T^2(0.01) = \frac{(n-1)(p-1)}{n-p+1} F_{0.99}(2,48)$$

$$= \frac{49 \times 2}{48} \times 5.08 = 10.37$$

显然 $T^2 > T^2(0.01)$,从而拒绝 H_0 即认为土壤各层间铅的含量有显著差异。

(二) 均值向量的子向量等于某已知向量的检验

设 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 是正态总体 $N_p(\boldsymbol{\mu}, \Sigma)(\Sigma > 0, n > p)$ 的样本,$\boldsymbol{\mu}$ 和 Σ 都未知,将 $\boldsymbol{\mu}$ 和 Σ 划分为

$$\boldsymbol{\mu}' = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

其中 $\boldsymbol{\mu}_1$ 为 r 维 $\boldsymbol{\mu}_2$ 为 $s = p - r$ 维. Σ_{11} 为 $r \times r$ 阶, Σ_{22} 为 $s \times s$ 阶, Σ_{12} 为 $r \times s$ 阶, $\Sigma_{12}' = \Sigma_{21}$. 我们要检验假设.

$$H_0: \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2^{(0)} \end{pmatrix}, \quad H_1: \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \neq \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2^{(0)} \end{pmatrix} \qquad (3.6)$$

此时

$\Theta = \{(\boldsymbol{\mu}, \Sigma): \boldsymbol{\mu}$ 为任意 p 维实向量, $\Sigma > 0\}$

$\Theta_0 = \{(\boldsymbol{\mu}, \Sigma): \boldsymbol{\mu}' = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2'), \boldsymbol{\mu}_2 = \boldsymbol{\mu}_2^{(0)}, \boldsymbol{\mu}_1$ 任意, $\Sigma > 0\}$

将样本作相应地划分

$$\boldsymbol{x}_i = \begin{pmatrix} \boldsymbol{x}_i^{(1)} \\ \boldsymbol{x}_i^{(2)} \end{pmatrix},\ \text{其中}\ \boldsymbol{x}_i^{(1)}\ \text{为}\ r\ \text{维},\ \boldsymbol{x}_i^{(2)}\ \text{为}\ s = p - r\ \text{维}$$

$$i = 1, 2, \cdots, n$$

样本平均及 $A = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'$ 的相应划分为

$$\bar{x} = \begin{pmatrix} \bar{x}^{(1)} \\ \bar{x}^{(2)} \end{pmatrix} \qquad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

由第二章多元正态变量边缘分布性质，可将 $x_i^{(2)} i = 1,2,\cdots,n$，看成 s 维正态总体 $N_s(\mu_2, \Sigma_{22})$ 的样本。利用我们熟知的似然比方法可导出似然比统计量为

$$\lambda^* = [1 + n(\bar{x}^{(2)} - \mu_2^{(0)})' A_{22}^{-1}(\bar{x}^{(2)} - \mu_2^{(0)})]^{-\frac{n}{2}}$$

它等价于统计量

$$\lambda = T^2 = (n-1)n(\bar{x}^{(2)} - \mu_2^{(0)})' A_{22}^{-1}(\bar{x}^{(2)} - \mu_2^{(0)}) \tag{3.7}$$

将(3.7)与(2.2)相比较便可得到如下结论。

定理 3.1 设

$$x_i = \begin{pmatrix} x_i^{(1)} \\ x_i^{(2)} \end{pmatrix} \; i = 1,2,\cdots,n$$

是从 $N_p(\mu, \Sigma)$ 中抽取的样本，($\Sigma > 0$，且未知，$n > p$) 其中 $x_i^{(1)}$ 为 r 维，$x_i^{(2)}$ 为 $s = p - r$ 维，

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

那么检验假设 (3.6) 的极大似然检验可简化为下述它的相应子向量的极大似然检验。

设 $x_i^{(2)}, i = 1,2,\cdots,n$ 是 $s = p - r$ 维正态总体 $N_s(\mu_2, \Sigma_{22})$ 的样本，相应检验假设为

$$H_0: \mu_2 = \mu_2^{(0)}, \quad H_1: \mu_2 \neq \mu_2^{(0)} \tag{3.8}$$

其拒绝域为

$$\begin{aligned} T^2 &= (n-1)n(\bar{x}^{(2)} - \mu_2^{(0)})' A_{22}^{-1}(\bar{x}^{(2)} - \mu_2^{(0)}) \\ &\geqslant T^2(\alpha) = \frac{(n-1)s}{n-s} F_{1-\alpha}(s, n-s) \end{aligned} \tag{3.9}$$

本节(一)中所讨论的对称性检验，实际上可化为子向量的检

验问题. 事实上,设 C 为 $(p-1)\times p$ 阶的对照阵,令

$$D=\begin{pmatrix}\mathbf{1}'_p\\ C\end{pmatrix}$$

显然 D 是满秩的 $p\times p$ 阶矩阵,设 $\boldsymbol{x}_i$, $i=1,2,\cdots,n$. 是总体 $N_p(\boldsymbol{\mu},\Sigma)$ 的样本,令 $\boldsymbol{y}_i=D\boldsymbol{x}_i$ $i=1,2,\cdots,n$, 则 $\boldsymbol{y}_i\sim N_p(\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\Sigma}})$ 且相互独立,其中

$$\tilde{\boldsymbol{\mu}}=D\boldsymbol{\mu}=\begin{pmatrix}\mathbf{1}'_p\boldsymbol{\mu}\\ C\boldsymbol{\mu}\end{pmatrix},\quad \tilde{\Sigma}=D\Sigma D'=\begin{pmatrix}\mathbf{1}'_p\Sigma\mathbf{1}_p & \mathbf{1}'_p\Sigma C'\\ C\Sigma\mathbf{1}_p & C\Sigma C'\end{pmatrix}$$

与对称性检验假设(3.4)比较,可看出(3.4)正是对正态总体 $N_p(\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\Sigma}})$ 的均值向量 $\tilde{\boldsymbol{\mu}}$ 的子向量 $C\boldsymbol{\mu}$ 是否等于零的检验.

(三) 均值向量的部分子向量已知,其它子向量等于某个常向量的检验

设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 是从正态总体 $N_p(\boldsymbol{\mu},\Sigma)(\Sigma>0$ 未知, $n>p)$ 中抽取的样本. 均值向量 $\boldsymbol{\mu}$ 和协方差阵 Σ 有如同(二)中的划分. 我们要检验假设

$$H_0:\begin{pmatrix}\boldsymbol{\mu}_1^{(0)}\\ \boldsymbol{\mu}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{\mu}_1^{(0)}\\ \boldsymbol{\mu}_2^{(0)}\end{pmatrix}=\boldsymbol{\mu}^{(0)},\quad H_1:\begin{pmatrix}\boldsymbol{\mu}_1^{(0)}\\ \boldsymbol{\mu}_2\end{pmatrix}\neq\begin{pmatrix}\boldsymbol{\mu}_1^{(0)}\\ \boldsymbol{\mu}_2^{(0)}\end{pmatrix}\tag{3.10}$$

其中 $\boldsymbol{\mu}_1^{(0)}$ 是 $\boldsymbol{\mu}$ 的 r 维子向量且为已知的, $\boldsymbol{\mu}_2^{(0)}$ 为指定的 $s=p-r$ 维常向量. 此时

$$\Theta=\left\{(\boldsymbol{\mu},\Sigma):\boldsymbol{\mu}_1=\boldsymbol{\mu}_1^{(0)},\ \boldsymbol{\mu}_2\text{ 任意},\ \boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}_1^{(0)}\\ \boldsymbol{\mu}_2\end{pmatrix},\ \Sigma>0\right\}$$

$$\Theta_0=\left\{(\boldsymbol{\mu},\Sigma):\boldsymbol{\mu}=\boldsymbol{\mu}^{(0)}=\begin{pmatrix}\boldsymbol{\mu}_1^{(0)}\\ \boldsymbol{\mu}_2^{(0)}\end{pmatrix},\ \Sigma>0\right\}\tag{3.11}$$

如果令 $\boldsymbol{y}=\boldsymbol{x}-\boldsymbol{\mu}^{(0)}$, 则 $\boldsymbol{y}\sim N_p(\tilde{\boldsymbol{\mu}},\Sigma)$,其中 $\tilde{\boldsymbol{\mu}}=\boldsymbol{\mu}-\boldsymbol{\mu}^{(0)}$. 那么(3.10)可转化为 $H_0:\tilde{\boldsymbol{\mu}}=0,H_1:\tilde{\boldsymbol{\mu}}\neq 0$. 因此为了后面的论证及叙述方便,不妨先假定 $\boldsymbol{\mu}^{(0)}=0$,此时(3.11)就可相应地改写为

$$\Theta'=\left\{(\boldsymbol{\mu},\Sigma):\boldsymbol{\mu}=\begin{pmatrix}0\\ \boldsymbol{\mu}_2\end{pmatrix},\boldsymbol{\mu}_2\text{ 任意},\ \Sigma>0\right\}$$

$$\Theta_0' = \{(\boldsymbol{\mu}, \Sigma): \boldsymbol{\mu} = 0, \Sigma > 0\} \tag{3.12}$$

利用似然比方法得到统计量

$$\lambda^* = \left(\frac{1 + n\bar{x}^{(1)'}A_{11}^{-1}\bar{x}^{(1)}}{1 + n\bar{x}'A^{-1}\bar{x}}\right)^{\frac{n}{2}} = \left(\frac{1}{\lambda}\right)^{\frac{n}{2}}$$

它等价于统计量

$$\lambda = \frac{1 + n\bar{x}'A^{-1}\bar{x}}{1 + n\bar{x}^{(1)'}A_{11}^{-1}\bar{x}^{(1)}} \tag{3.13}$$

下面我们要导出 $\lambda - 1$ 在给定 $\bar{x}^{(1)}$ 及 A_{11} 之下的条件分布，相应之下的检验称为条件检验。

根据第一章第二节分块矩阵的求逆证明中有

$$A^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}\begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{2\cdot1}^{-1} \end{pmatrix}\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}$$

其中 $A_{2\cdot1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$，对于 $\bar{x} = \begin{pmatrix} \bar{x}^{(1)} \\ \bar{x}^{(2)} \end{pmatrix}$ 有

$$\bar{x}'A^{-1}\bar{x} = (\bar{x}^{(2)} - A_{21}A_{11}^{-1}\bar{x}^{(1)})'A_{2\cdot1}^{-1}(\bar{x}^{(2)} - A_{21}A_{11}^{-1}\bar{x}^{(1)}) + \bar{x}^{(1)'}A_{11}^{-1}\bar{x}^{(1)}$$

从而

$$\lambda - 1 = \frac{n(\bar{x}^{(2)} - A_{21}A_{11}^{-1}\bar{x}^{(1)})'A_{2\cdot1}^{-1}(\bar{x}^{(2)} - A_{21}A_{11}^{-1}\bar{x}^{(1)})}{1 + n\bar{x}^{(1)'}A_{11}^{-1}\bar{x}^{(1)}} \tag{3.14}$$

由于参数空间 Θ' 中子向量 $\boldsymbol{\mu}_1 = 0$，再根据正态向量的条件分布以及第三章中的定理 2.6，有

i) 给定 $\bar{x}^{(1)}$ 时，$\sqrt{n}\,\bar{x}^{(2)} \sim N_s(\sqrt{n}(\boldsymbol{\mu}_2 + \beta\bar{x}^{(1)}), \Sigma_{2\cdot1})$，其中 $\beta = \Sigma_{21}\Sigma_{11}^{-1}$，$\Sigma_{2\cdot1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$。

ii) 给定 A_{11} 时，$A_{21} \sim N_{s,r}(\beta A_{11}, \Sigma_{2\cdot1}, A_{11})$。从而给定 $\bar{x}^{(1)}$ 和 A_{11} 后

$$\sqrt{n}\,A_{21}A_{11}^{-1}\bar{x}^{(1)} \sim N_s(\sqrt{n}\,\beta\bar{x}^{(1)}, n\Sigma_{2\cdot1}(\bar{x}^{(1)'}A_{11}^{-1}\bar{x}^{(1)}))$$

iii) $A_{2\cdot1} \sim W_s(n - r - 1, \Sigma_{2\cdot1})$，且与 (A_{21}, A_{11}) 相互独立，由于 $\bar{x}$ 与 A 相互独立，从而给定 $\bar{x}^{(1)}$ 和 A_{11} 后

$$\sqrt{n}(\bar{x}^{(2)} - A_{21}A_{11}^{-1}\bar{x}^{(1)}) \sim N_s(\sqrt{n}\,\mu_2,$$
$$(1 + n\bar{x}^{(1)'}A_{11}^{-1}\bar{x}^{(1)})\Sigma_{2\cdot 1})$$

即

$$\frac{\sqrt{n}(\bar{x}^{(2)} - A_{21}A_{11}^{-1}\bar{x}^{(1)})}{(1 + n\bar{x}^{(1)'}A_{11}^{-1}\bar{x}^{(1)})^{\frac{1}{2}}} \sim N_s\left(\frac{\sqrt{n}\,\mu_2}{(1 + n\bar{x}^{(1)'}A_{11}^{-1}\bar{x}^{(1)})^{\frac{1}{2}}},\ \Sigma_{2\cdot 1}\right)$$

于是按第三章第三节 Hotelling 的 T^2 统计量的定义，给定 $\bar{x}^{(1)}$ 和 A_{11} 后，$(n-r-1)(\lambda-1)$ 正是自由度为 s 和 $n-r-1-s+1=n-p$ 的 T^2 统计量. 由第三章定理 3.1

$$\frac{n-p}{s\cdot(n-r-1)}T^2 = \frac{n-p}{s}(\lambda-1)$$
$$= \frac{n-p}{s}\,\frac{n(\bar{x}^{(2)} - A_{21}A_{11}^{-1}\bar{x}^{(1)})'A_{2\cdot 1}^{-1}(\bar{x}^{(2)} - A_{21}A_{11}^{-1}\bar{x}^{(1)})}{1 + n\bar{x}^{(1)'}A_{11}^{-1}\bar{x}^{(1)}}$$
$$\sim F(s, n-p, \Omega)$$

其中非中心参数

$$\Omega = \frac{n\mu_2'\Sigma_{2\cdot 1}^{-1}\mu_2}{1 + n\bar{x}^{(1)'}A_{11}^{-1}\bar{x}^{(1)}}$$

在原假设 $H_0:\mu=0$ 为真时，给定 $\bar{x}^{(1)}$ 和 A_{11} 后，上述的 F 分布就成为中心的，即

$$\frac{n-p}{s}(\lambda-1) \sim F(s, n-p) \tag{3.15}$$

关于检验假设(3.10)也可得到类似结果，于是有如下定理。

定理 3.2 设 $x_1,\cdots,x_n$ 是取自正态总体 $N_p(\mu,\Sigma)$ 容量为 n 的样本 $(\Sigma>0$ 且未知，$n>p)$，令 $\mu'=(\mu_1',\mu_2')$ 其中 μ_1 是 r 维，μ_2 是 $s=p-r$ 维对于检验假设

$$H_0:\mu=0,\ H_1:\mu_1=0,\ \mu_2\neq 0$$

其极大似然条件检验的拒绝域是(当给定 $\bar{x}^{(1)}$ 和 A_{11} 时)：

$$\frac{n-p}{s}(\lambda-1) \geqslant F_{1-\alpha}(s, n-p) \tag{3.16}$$

其中 λ 为(3.13)式所示，α 为显著水平. 而相应于检验假设(3.10)

的极大似然条件检验的拒绝域是(当给定 $\bar{x}^{(1)}$ 和 A_{11} 时):

$$\frac{n-p}{s}(\lambda-1) \geqslant F_{1-\alpha}(s, n-p) \tag{3.17}$$

其中

$$\lambda = \frac{1+n(\bar{x}-\mu^{(0)})' A^{-1}(\bar{x}-\mu^{(0)})}{1+n(\bar{x}^{(1)}-\mu_1^{(0)})' A_{11}^{-1}(\bar{x}^{(1)}-\mu_1^{(0)})} \tag{3.18}$$

在实用中,常利用下述的记号,令

$$T_p^2 = n(\bar{x}-\mu^{(0)})' S^{-1}(\bar{x}-\mu^{(0)})$$
$$T_s^2 = n(\bar{x}^{(1)}-\mu_1^{(0)}) S_{11}^{-1}(\bar{x}^{(1)}-\mu_1^{(0)})$$

其中

$$S = \frac{1}{n-1} A = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

则

$$\lambda - 1 = \frac{T_p^2 - T_s^2}{(n-1)+T_s^2} \tag{3.19}$$

当拒绝了 H_0 后可以写出 μ_2 的 $1-\alpha$ 置信区域是

$$\frac{n[\bar{x}^{(2)}-A_{21}A_{11}^{-1}(\bar{x}^{(1)}-\mu_1^{(0)})-\mu_2]' A_{22\cdot 1}^{-1} \times [\bar{x}^{(2)}-A_{21}A_{11}^{-1}(\bar{x}^{(1)}-\mu_1^{(0)})-\mu_2]}{1+n(\bar{x}^{(1)}-\mu_1^{(0)})' A_{11}^{-1}(\bar{x}^{(1)}-\mu_1^{(0)})}$$
$$\leqslant \frac{s}{n-p} F_{1-\alpha}(s, n-p) \tag{3.20}$$

例 3.2 任意选择 10 个学生,对他们两次考试成绩进行分析.其中一次事先未经过模拟考试直接进行(记为 I),另一次事先经过模拟考试再进行(记为 II)数据见表 3.1. 希望通过他们的成绩了解模拟考试是否有一定的效果. 注意到模拟是否有效果还与学生的基础有关,因此假定已知所有学生的平时成绩是 50 分. 对于此问题可作如下的检验假设:

$$H_0: \mu_1 = 50,\ \mu_2 = 50,\ H_1: \mu_1 = 50,\ \mu_2 \neq 50$$

经计算求得样本均值向量、协方差阵及相关矩阵为

表 3.1 10个学生两次考试成绩

考试＼学生	1	2	3	4	5	6	7	8	9	10
I	70	60	65	50	43	40	80	45	30	40
II	75	58	70	55	48	41	78	65	55	50

$$\bar{x} = \begin{pmatrix} \bar{x}^{(1)} \\ \bar{x}^{(2)} \end{pmatrix} = \begin{pmatrix} 52.3 \\ 59.5 \end{pmatrix} \quad S = \begin{pmatrix} 250 & 159 \\ 159 & 148 \end{pmatrix} \quad R = \begin{pmatrix} 1 & 0.83 \\ 0.83 & 1 \end{pmatrix}$$

由 $r = 0.83$ 知二次考试成绩相关程度较高，这是符合实际的。

$$T_r^2 = n(\bar{x}^{(1)} - \mu_1^{(0)})S_{11}^{-1}(\bar{x}^{(1)} - \mu_1^{(0)})$$

$$= 10\,(52.3 - 50)\cdot\frac{1}{250}\cdot(52.3 - 50) = 0.21$$

$$T_p^2 = n(\bar{x} - \mu^{(0)})'S^{-1}(\bar{x} - \mu^{(0)}) = 14.1$$

$$\lambda - 1 = \frac{T_p^2 - T_r^2}{(n-1) + T_r^2} = \frac{13.89}{9.21} = 1.51$$

$$\frac{n-p}{s}(\lambda - 1) = 8 \times 1.51 = 12.06$$

查 F 表得 $F_{0.95}(1,8) = 5.32$，显然 $12.06 > 5.32$，所以拒绝 H_0，即模拟考试有一定效果。注意到

$$A = 9 \cdot S = \begin{pmatrix} 2250 & 1431 \\ 1431 & 1332 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

还可由(3.20)得到 μ_2 的 0.95 置信区间为

$$52.80 \leqslant \mu_2 \leqslant 63.48.$$

(四) 两个总体的均值向量的子向量相等的检验

设 $x_1, \cdots, x_{n_1}$ 是正态总体 $N_p(\mu, \Sigma)$ 的容量为 n_1 的样本，$y_1, \cdots, y_{n_2}$ 是正态总体 $N_p(\nu, \Sigma)$ 的容量为 n_2 的样本，其中 $\Sigma > 0$ 且未知，两样本间相互独立。对 μ、ν 及 Σ 作如下的划分

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\nu} = \begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

其中 $\boldsymbol{\mu}_1$，$\boldsymbol{\nu}_1$ 分别是 $\boldsymbol{\mu}$ 和 $\boldsymbol{\nu}$ 的 r 维子向量，$\boldsymbol{\mu}_2$ 和 $\boldsymbol{\nu}_2$ 分别是它们的 $p-r=s$ 维子向量，Σ_{ij} 也有相应的解释。我们要检验的假设是

$$H_0: \boldsymbol{\mu} = \boldsymbol{\nu}, \quad H_1: \boldsymbol{\mu}_1 = \boldsymbol{\nu}_1, \boldsymbol{\mu}_2 \neq \boldsymbol{\nu}_2 \tag{3.21}$$

检验(3.21)的思想方法与(三)完全一样，读者可自行推导。在这里我们直接给出其检验统计量：

$$\lambda = \frac{\left[1 + \frac{n_1 n_2}{n_1 + n_2}(\bar{x} - \bar{y})' A^{-1}(\bar{x} - \bar{y})\right]}{\left[1 + \frac{n_1 n_2}{n_1 + n_2}(\bar{x}^{(1)} - \bar{y}^{(1)})' A_{11}^{-1}(\bar{x}^{(1)} - \bar{y}^{(1)})\right]} \tag{3.22}$$

如果记

$$T_p^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{x} - \bar{y})' S^{-1}(\bar{x} - \bar{y})$$

$$T_r^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{x}^{(1)} - \bar{y}^{(1)})' S_{11}^{-1}(\bar{x}^{(1)} - \bar{y}^{(1)})$$

则

$$\lambda - 1 = \frac{T_p^2 - T_r^2}{(n_1 + n_2 - 2) + T_r^2} \tag{3.23}$$

其中

$$\bar{x} = \begin{pmatrix} \bar{x}^{(1)} \\ \bar{x}^{(2)} \end{pmatrix} \quad \bar{y} = \begin{pmatrix} \bar{y}^{(1)} \\ \bar{y}^{(2)} \end{pmatrix} \quad S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \quad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

而

$$S = \frac{A}{n_1 + n_2 - 2}$$

在 H_0 为真时，给定 $\bar{x}^{(1)}$、$\bar{y}^{(1)}$ 及 A_{11} 的条件下有

$$\frac{n_1 + n_2 - p - 1}{s}(\lambda - 1) \sim F(s, n_1 + n_2 - p - 1) \tag{3.24}$$

利用(3.24)就可以对(3.21)进行条件检验。

例 3.3　在两周的时间中，把 16 条鱼分成二组分别用通常标

准饲料和改进后的饲料喂养．观察其体长的增长，数据列于表 3.2．由于最初鱼是随机选取分组，故可认为这两组鱼初始时平均长度是相同的，我们希望了解改进后的饲料是否有显著成效．即我们要检验假设

$$H_0: \mu_i = \nu_i (i = 1,2,3),\ H_1: \mu_1 = \nu_1,\ \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} \neq \begin{pmatrix} \nu_2 \\ \nu_3 \end{pmatrix}$$

表 3.2　两组鱼喂不同饲料的体长增长数据

	鱼编号	1	2	3	4	5	6	7	8	总体均值
标准饲料	起始长度	12.3	12.1	12.8	12.0	12.1	11.8	12.7	12.5	μ_1
	一周后	2.5	2.2	2.9	2.1	2.2	1.9	2.9	2.7	μ_2
	二周后	2.9	2.5	3.0	2.2	2.4	2.0	3.3	3.0	μ_3
改进饲料	起始长度	12.0	11.8	12.7	12.4	12.1	12.0	11.7	12.2	ν_1
	一周后	2.3	2.0	3.1	2.8	2.5	2.2	2.0	2.5	ν_2
	二周后	2.7	2.4	3.6	3.2	2.8	2.7	2.4	3.0	ν_3

经计算得到

$$\bar{\boldsymbol{x}} = (12.288, 2.425, 2.663)'$$

$$\bar{\boldsymbol{y}} = (12.113, 2.425, 2.850)'$$

$$S = \frac{A}{14} = \begin{pmatrix} 0.114 & 0.128 & 0.140 \\ 0.128 & 0.146 & 0.161 \\ 0.140 & 0.161 & 0.186 \end{pmatrix}$$

此例中 $n_1 + n_2 - 2 = 8, p = 3, r = 1,\ s = 3 - 1 = 2$．再计算 $T_p^2 = 22.155, T_r^2 = 1.22$．代入(3.24)得

$$\frac{n_1 + n_2 - p - 1}{s}(\lambda - 1) = \frac{12}{2} \cdot 1.373 = 8.238$$

查 F 表得 $F_{0.95}(2,12) = 3.89$．显然 $8.238 > 3.89$，从而拒绝 H_0，即改进饲料对鱼体的增长有显著成效．

*§ 5.4　总体均值的大样本推断

前面所讲的总体均值向量的检验，都是在总体为正态总体的前题下进行的。如何对非正态总体的均值向量作出推断呢？与一元统计相类似，我们要利用第四章中所介绍的多元中心极限定理和大数定律来解决。在本节中我们仅介绍单总体情况以及实际生产中经常遇到的多项分布总体的参数推断。

（一）单总体均值向量的大样本检验

设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 为来自均值向量为 $\boldsymbol{\mu}$，协方差矩阵为 $\Sigma(\Sigma>0$，未知)的总体的样本，我们要检验假设

$$H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0,\quad H_1:\boldsymbol{\mu}\neq\boldsymbol{\mu}_0 \tag{4.1}$$

记

$$\bar{\boldsymbol{x}}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i,\ S=\frac{A}{n-1}=\frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'$$

采用的检验统计量为

$$\begin{aligned}\chi^2&=n(n-1)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'A^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)\\&=n(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'S^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)\end{aligned}$$

其结构与正态总体的情况一样，但现在总体为非正态总体，上述统计的精确分布是无法知道的。根据第四章 § 4.1 的讨论知道：只要 n 充分大或者说 $n-p$ 很大时，$n(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'S^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)$ 就近似地服从 $\chi^2(p)$ 分布。对于显著水平 α，若 $n(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'S^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)\geqslant\chi^2_{1-\alpha}(p)$，则拒绝原假设 H_0。这种检验的显著水平近似地等于 α。

我们把上述检验与 § 5.1 中的（一）相比较可看到两个检验的统计量有相同结构，但临界值不一样，一个是 $\chi^2_{1-\alpha}(p)$ 另一个是 $\dfrac{p(n-1)}{n-p}F_{1-\alpha}(p,n-p)$。更进一步地研究表明，如果 $n(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'S^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)$ 与 χ^2 分布近似程度比较好时，这两个检验给出了

本质上相同的结果。这是因为当 $(n-p)$ 较大时，$\frac{p(n-1)}{n-p}\cdot F_{1-\alpha}(p,n-p)$ 与 $\chi^2_{1-\alpha}(p)$ 近似相等。

当拒绝 H_0 时，μ 的 $1-\alpha$ 置信区域近似地为

$$n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu)<\chi^2_{1-\alpha}(p) \tag{4.2}$$

当 $n-p$ 很大时，对于一切 $l\neq 0$ 的 p 维实向量，$l'\mu$ 的 $1-\alpha$ 同时置信区间近似地为

$$\left[l'\bar{x}\pm\sqrt{\chi^2_{1-\alpha}(p)}\cdot\sqrt{\frac{l'Sl}{n}}\right] \tag{4.3}$$

特别取 $l=(0,\cdots,0,1,0,\cdots,0)'$，其中第 i 个分量为 1 其余为 0，则 $\mu_i,i=1,2,\cdots,p$ 的 $1-\alpha$ 同时置信区间近似地为

$$\left[\bar{x}_i\pm\sqrt{\chi^2_{1-\alpha}(p)}\cdot\sqrt{\frac{s_{ii}}{n}}\right] \tag{4.4}$$

其中 $\bar{x}_i$ 为 $\bar{x}$ 的第 i 个分量，s_{ii} 为 S 的 (ii) 元素，特别取 $l=(0,\cdots,0,1,0,\cdots,0,-1,0,\cdots,0)'$，其中第 i 个分量为 1，第 k 个分量为 $-1(i<k,k=2,3,\cdots,p)$，则对所有的 $(\mu_i-\mu_k)$ 的 $1-\alpha$ 同时置信区间近似地为

$$\left[(\bar{x}_i-\bar{x}_k)\pm\sqrt{\chi^2_{1-\alpha}(p)}\cdot\sqrt{\frac{s_{ii}-2s_{ik}+s_{kk}}{n}}\right] \tag{4.5}$$

其中 $\bar{x}_i,\bar{x}_k$ 为 $\bar{x}$ 的第 i 分量和第 k 分量，s_{ik} 为 S 的(i,k)元素。上述结论的推导可参看[26]。

（二） 多项分布总体的参数推断

在实际问题中，经常会遇到这样一种总体，它的每一个个体要么具有某属性，要么不具有该属性。例如所生产的产品中，每一个产品不是次品便是正品。这种总体由两个不相同的类组成。个体 X_i 是一个服从两点分布的随机变量：

$$X_i=\begin{cases}1 & \text{具有某属性}\\ 0 & \text{不具某属性}\end{cases}$$

设 p 为具有某属性的个体在总体中所占的比例，显然，X_i 的概率

分布为

$$P(X_i = x_i) = p^{x_i}(1-p)^{1-x_i} \quad x_i = 0 \text{ 或 } 1$$

我们感兴趣的是对参数 p 的推断，这在一元统计中已有详尽的讨论。

如果总体中的个体分为 $m+1$ 个不相同的类，且个体属于第 k 个类（$k=1,2,\cdots,m+1$）所占的比例为 $p_k, k=1,2,\cdots,m+1$，且 $p_{m+1}=1-\sum_{k=1}^{m} p_k$。令来自第 k 类的个体对应一个 $(m+1)\times 1$ 向量：$(0,0,\cdots,0,1,0,\cdots,0)'$，其中 1 出现在第 k 个分量的位置上。众所周知由这种个体组成的总体其概率分布为 $m+1$ 项分布（多项分布），写成向量就具有如下结构形式：

类别	1	2	$\cdots$	k	$\cdots$	m	$m+1$
观测向量	$\begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix}$	$\begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix}$	$\cdots$	$\begin{pmatrix}0\\\vdots\\0\\1\\0\\\vdots\\0\end{pmatrix}$	$\cdots$	$\begin{pmatrix}0\\0\\\vdots\\0\\1\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\\\vdots\\0\\1\end{pmatrix}$
概率	p_1	p_2	$\cdots$	p_k	$\cdots$	p_m	$p_{m+1}=1-\sum_{k=1}^{m} p_k$

(4.6)

设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 为来自由 $m+1$ 个不同类组成的总体的样本。若 $\boldsymbol{x}_i$ 来自第 k 类，则 $\boldsymbol{x}_i$ 所对应的向量的第 k 个分量 $x_{ki}=1$，其余分量 $x_{ji}=0$（$j\neq k$）。而 $\boldsymbol{x}_i$ 来自第 k 类的概率为 p_k，所以有 $P(x_{ki}=1)=p_k$，于是 $\mu_k=E(x_{ki})=p_k, \sigma_{kk}=\operatorname{var}(x_{ki})=p_k(1-p_k)$。因为当 $j\neq k$ 时 x_{ki} 与 x_{ji} 不能同时为 1，所以 $\boldsymbol{x}_i$ 的第 k 个分量 x_{ki} 与第 j 个分量 x_{ji} 的协方差为

$$\sigma_{kj}=\operatorname{cov}(x_{ki},x_{ji})=-p_k p_j$$

从而 $\boldsymbol{x}_i$ 的方差协方差阵为

$$\Sigma = (\sigma_{kj})_{(m+1)\times(m+1)} \text{ 其中 } \sigma_{kj} = \begin{cases} p_k(1-p_k) & \text{当 } k = j \\ -p_k p_j & \text{当 } k \neq j \end{cases} \tag{4.7}$$

按照结构(4.6),样本 $x_i, i = 1,2,\cdots,n$ 的样本均值向量为

$$\hat{p} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = (\hat{p}_1, \hat{p}_2, \cdots, \hat{p}_{m+1})' \tag{4.8}$$

其中

$$\hat{p}_k = \frac{1}{n}\sum_{i=1}^{n} x_{ki}, \quad x_{ki} = \begin{cases} 1 & \text{当 } x_i \text{ 属于第 } k \text{ 类} \\ 0 & \text{否则} \end{cases}$$

$\hat{p}_k$ 表示了样本中属于第 k 类的个体所占的比例. 显然有

$$p = E(\bar{x}) = (p_1, p_2, \cdots, p_{m+1})', \operatorname{cov}(\bar{x}, \bar{x}) = \frac{1}{n}\Sigma$$

其中 Σ 为 (4.7) 所示. 由第四章的多元中心极限定理知道 $\sqrt{n}\cdot(\bar{x} - E\bar{x}) = \sqrt{n}(\hat{p} - p)$ 的极限分布为 $N_{m+1}(0,\Sigma)$,即当 n 较大时

$$\sqrt{n}(\hat{p} - p) \text{近似服从} N_{m+1}(0,\Sigma) \tag{4.9}$$

如果 σ_{kk}, σ_{kj} 分别用它们的估计 $\hat{\sigma}_{kk} = \hat{p}_k(1 - \hat{p}_k)$, $\hat{\sigma}_{kj} = -\hat{p}_k\hat{p}_j$ $(k \neq j)$ 来代替,(4.9) 式仍近似成立. 即 $\sqrt{n}(\hat{p} - p)$ 近似服从 $N_{m+1}(0,\hat{\Sigma})$,其中 $\hat{\Sigma}$ 的元素为 $\hat{\sigma}_{kk}$ 和 $\hat{\sigma}_{kj}$. 注意到 Σ 和 $\hat{\Sigma}$ 的秩都是 m. 它们的逆矩阵不存在，但我们仍然可以近似构造所有线性组合 $l'p$ 的 $1-\alpha$ 同时置信区间(参看 [26]):

$$\left[l'\hat{p} \pm \sqrt{\chi^2_{1-\alpha}(m)} \cdot \sqrt{\frac{l'\hat{\Sigma}l}{n}}\right] \tag{4.10}$$

其中 $\hat{p}$ 为 (4.8) 所示，根据实际问题的需要可选取不同的 l. 在应用中，充分大的 n 可理解为 n 满足 $n\hat{p}_k \geqslant 20$, $k = 1, 2, \cdots, m+1$.

例 4.1 为了研究洗衣机在某地区的销售情况,对 A、B、C 三个主要厂家的产品在该地区的销售比例进行调查，每个被调查者回答下述问题：你购买的洗衣机是 A、B、C 中的哪一家? 或者是

其它厂家。设 A、B、C 三个厂家的销售比例为 p_1,p_2,p_3，而其它厂家的销售比例为 $p_4=1-\sum_{k=1}^{3}p_k$。有答复的户数为550户，调查的数据列在表4.1中。

表 4.1　洗衣机销售情况调查数据

厂　家	A	B	C	其　它
购买户数	187	215	88	60
总体比例	p_1	p_2	p_3	$p_4=1-\sum_{k=1}^{3}p_k$
样本比例	0.34	0.39	0.16	0.11

由表中样本比例 $\hat{p}_k$，$k=1,2,3,4$，计算得

$$\hat{\Sigma}=\begin{pmatrix} 0.2244 & -0.1326 & -0.0544 & -0.0374 \\ -0.1326 & 0.2379 & -0.0624 & -0.0429 \\ -0.0544 & -0.0624 & 0.1344 & -0.0176 \\ -0.0374 & -0.0429 & -0.0176 & 0.0979 \end{pmatrix}$$

选取 $\boldsymbol{l}=(1,0,0,0)'$，$\boldsymbol{l}=(0,1,0,0)'$，$\boldsymbol{l}=(0,0,1,0)'$ 及 $\boldsymbol{l}=(0,0,0,1)'$，对 $\alpha=0.05$ 查 χ^2 表得 $\chi^2_{0.95}(3)=7.815$。根据(4.10)计算得 p_1,p_2,p_3,p_4 的0.95同时置信区间为

$$0.28\leqslant p_1\leqslant 0.40,\quad 0.33\leqslant p_2\leqslant 0.45,$$

$$0.11\leqslant p_3\leqslant 0.21,\quad 0.07\leqslant p_4\leqslant 0.15.$$

若选取 $\boldsymbol{l}=(1,-1,0,0)'$，则 $\boldsymbol{l}'\boldsymbol{p}=p_1-p_2$。此时是把 A、B 两厂进行比较。经计算得 p_1-p_2 的0.95置信区间为 $-0.16\leqslant p_1-p_2\leqslant 0.06$，这个区间包含了0，可以认为 A、B 两厂在该地区的销售比例无显著差异。

§5.5　协方差阵的检验

本节是讨论对多元正态总体的协方差阵的检验，检验的统计

量都是利用似然比方法导出的，其分布绝大部分都是采用近似分布。

（一） 协方差阵等于已知矩阵的检验

设 $x_1, \cdots, x_n$ 是来自正态总体 $N_p(\mu, \Sigma)(\Sigma > 0, n > p)$ 的样本，μ 和 Σ 都是未知，我们要检验的假设是

$$H_0: \Sigma = \Sigma_0;\quad H_1: \Sigma \neq \Sigma_0 \tag{5.1}$$

其中 Σ_0 为给定的正定矩阵.

因为 Σ_0 正定，存在非奇异矩阵 D 使得 $D\Sigma_0 D' = I$. 令 $y_i = Dx_i$, $i = 1, 2, \cdots, n$, 则

$$y_i \sim N_p(D\mu, D\Sigma D') = N_p(\mu^*, \Sigma^*).$$

这时相应于(5.1)的假设等价于

$$H_0: \Sigma^* = I;\quad H_1: \Sigma^* \neq I \tag{5.2}$$

定理 5.1 设 $x_i, i = 1, 2, \cdots, n$ 为来自总体 $N_p(\mu, \Sigma)$ 的样本，检验假设(5.1)的似然比统计量为

$$\lambda = \left(\frac{e}{n}\right)^{\frac{np}{2}} |\Sigma_0^{-1}A|^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\Sigma_0^{-1}A\right\} \tag{5.3}$$

其中

$$A = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})',\quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

证明 由前所述，令 $y_i = Dx_i$, D 为非奇异阵，检验(5.1)等价于检验 (5.2). 此时似然函数为

$$L(\mu^*, \Sigma^*) = (2\pi)^{-\frac{np}{2}} |\Sigma^*|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\Sigma^{*-1}\right.\right.$$

$$\left.\left.\cdot\left(\sum_{i=1}^{n}(y_i - \mu^*)(y_i - \mu^*)'\right)\right]\right\}$$

参数空间 $\Theta = \{(\mu^*, \Sigma^*): \mu^*$ 为任意 p 维实向量, $\Sigma^* > 0\}$, 在 H_0 之下，子集 $\Theta_0 = \{(\mu^*, \Sigma^*): \mu^*$ 为任意 p 维实向量, $\Sigma^* = I\}$.

$$\max_{(\mu^*,\Sigma^*)\in\Theta} L(\mu^*,\Sigma^*) = \left(\frac{2\pi}{n}\right)^{-\frac{np}{2}} |B|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}np\right\}$$

其中

$$B = \sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})', \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

而

$$\max_{(\mu^*,\Sigma)\in\Theta_0} L(\mu^*,\Sigma^*) = (2\pi)^{-\frac{np}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr} B\right\}$$

于是似然比统计量

$$\lambda = \left(\frac{e}{n}\right)^{\frac{np}{2}} |B|^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr} B\right\} \tag{5.4}$$

由于 $D\Sigma_0 D' = I$，即 $D'D = \Sigma_0^{-1}$，又 $y_i = Dx_i$，所以

$$\begin{aligned} B &= \sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})' \\ &= D\left[\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'\right]D' = DAD' \end{aligned}$$

$$|B| = |DAD'| = |\Sigma_0^{-1}A|$$

$$\operatorname{tr} B = \operatorname{tr}(DAD') = \operatorname{tr}(\Sigma_0^{-1}A)$$

将它们代入(5.4)便得到(5.3)。 ∎

由于以 λ 为基础的似然比检验不是无偏的，将 λ 中的 n 换成 $n-1$ 后得到的似然比检验可证明它是无偏的。λ 统计量等价于统计量 $-2\ln\lambda$。

定理 5.2 在 H_0 为真时，$-2\ln\lambda$ 的极限分布是自由度为 $\frac{1}{2}p(p+1)$ 的 χ^2 分布。

该定理的证明见[7]。

检验的具体步骤如下：

i) 根据样本值 x_i，$i = 1,2,\cdots,n$。由(5.3)计算出 λ 值(注

意 λ 中的 n 要用 $n-1$ 代替)；

ii) 由给定的显著水平 α，查自由度为 $f=\frac{1}{2}p(p+1)$ 的 χ^2 分布表得临界值 $\chi^2_{1-\alpha}(f)$；

iii) 若 $-2\ln\lambda \geqslant \chi^2_{1-\alpha}(f)$ 或 $\lambda \leqslant e^{-\frac{1}{2}\chi^2_{1-\alpha}(f)}$，则拒绝原假设 $H_0: \Sigma=\Sigma_0$.

例 5.1 在服装标准制定中，协方差矩阵是重要的参数．从长期的调查中已知北京市成年女子身高 x_1，胸围 x_2 和腰围 x_3 的协方差阵为

$$\Sigma_0=\begin{pmatrix} 29.57 & & \\ 3.92 & 39.05 & \\ 1.76 & 39.19 & 63.07 \end{pmatrix}.$$

为了验证，近期抽取 100 位成年女子，经测量并计算得这三个变量的样本协方差矩阵为

$$S=\frac{1}{99}A=\begin{pmatrix} 22.12 & & \\ 2.98 & 32.72 & \\ 0.46 & 30.39 & 50.98 \end{pmatrix}$$

我们希望知道近期的成年女子的身高、胸围和腰围的协方差阵是否与原来的 Σ_0 一致．通过计算有

$$\Sigma_0^{-1}=\begin{pmatrix} 0.03450 & & \\ -0.0066 & 0.0693 & \\ 0.00316 & -0.0429 & 0.0424 \end{pmatrix}$$

$$\mathrm{tr}(\Sigma_0^{-1}S)=2.5474$$

$$|\Sigma_0|=26862.84, \quad \ln|\Sigma_0|=10.198$$

$$|S|=16092.31 \qquad \ln|S|=9.686$$

将它们代进 $-2\ln\lambda$ 得

$$\begin{aligned} -2\ln\lambda &= (n-1)[\ln|\Sigma_0|-\ln|S|-p+\mathrm{tr}(\Sigma_0^{-1}S)] \\ &= 99\cdot[10.198-9.686-3+2.5474]=5.88 \end{aligned}$$

对于 $\alpha=0.05$，查 χ^2 表得 $\chi^2_{0.95}(6)=12.59$，显然，$-2\ln\lambda < \chi^2_{0.95}(6)$，从而不能拒绝 H_0.

（二） 球性检验

设 $x_i, i=1,2,\cdots,n$. 为总体 $N_p(\mu,\Sigma)$ 的样本 $(n>p)$，其中 μ 和 $\Sigma(\Sigma>0)$ 都未知，要检验假设

$$H_0: \Sigma=\sigma^2\Sigma_0,\quad H_1: \Sigma\neq\sigma^2\Sigma_0 \tag{5.5}$$

其中 Σ_0 为给定的正定矩阵，σ^2 是一个未指定的正常数。此假设的含义为总体协方差 Σ 与给定的正定矩阵 Σ_0 是否成比例。

类似于(一)的讨论，令 $y_i=Dx_i$，$i=1,2,\cdots,n$. 其中 D 为满足 $D\Sigma_0D'=I$ 的非奇异阵，则 $y_i, i=1,2,\cdots,n$. 为来自 $N_p(D\mu,D\Sigma D')=N_p(\mu^*,\Sigma^*)$ 的样本，(5.5)的假设化为

$$H_0: \Sigma^*=\sigma^2I,\quad H_1: \Sigma^*\neq\sigma^2I \tag{5.6}$$

(5.6) 的 H_0 为真时，说明随机向量的各个分量之间相互独立且有相同的未知方差。这相当于正态密度的等值面由一般的椭球面 $(y-\mu^*)'\Sigma^{*-1}(y-\mu^*)=c$（$c$ 为常数）转化为一个球面 $(y-\mu^*)'(y-\mu^*)=c\sigma^2$，故称为球性检验。如果 σ^2 已知且等于1，则就是(一)所讨论的问题。

利用似然比方法及(一)中所用的记号，我们有以下定理。

定理 5.3 设 $x_i, i=1,2,\cdots,n$. 为总体 $N_p(\mu,\Sigma)$ 的样本，$(n>p)$，其中 μ 和 $\Sigma(\Sigma>0)$ 都未知，检验假设(5.5)的似然比统计量为

$$\lambda^*=\frac{|\Sigma_0^{-1}A|^{\frac{n}{2}}}{\left[\frac{1}{p}\operatorname{tr}(\Sigma_0^{-1}A)\right]^{\frac{np}{2}}} \tag{5.7}$$

证明留给读者完成。

在实用中，我们取等价统计量

$$\lambda=(\lambda^*)^{\frac{2}{n}}=\frac{p^p|\Sigma_0^{-1}A|}{[\operatorname{tr}(\Sigma_0^{-1}A)]^p} \tag{5.8}$$

Mauchly 在 1940 年导出了 λ 的各阶矩，Davis 在 1971 年获得了修改了的 λ（在 λ 中的 n 代以 $n-1$）的近似分布并证明了如下结

论(见[7])。

定理 5.4 在(5.5)中的 H_0 为真时，当 $n \to \infty$ 时 $-(n-1)\cdot \beta \ln \lambda$ 渐近于自由度为 $\frac{1}{2}p(p+1)-1$ 的 χ^2 分布，其中 $\beta = 1 - \frac{2p^2+p+2}{6p(n-1)}$。

(三) 均值向量和协方差矩阵等于给定向量和矩阵的检验

设 $x_i, i=1,2,\cdots,n$ 是从正态总体 $N_p(\mu,\Sigma)$ 中抽取的样本 $(n>p)$，其中 μ 及 $\Sigma(\Sigma>0)$ 都是未知的，要检验的假设是

$$
\begin{aligned}
&H_0: \mu=\mu_0,\ \Sigma=\Sigma_0,\\
&H_1: \mu=\mu_0, \Sigma=\Sigma_0 \text{ 至少有一个等式不成立}
\end{aligned} \tag{5.9}
$$

其中 μ_0 是给定的向量，Σ_0 是给定的正定矩阵。由于 Σ_0 正定，存在非奇异矩阵 D 满足 $D\Sigma_0 D' = I$，作变换 $y_i = D(x_i-\mu_0)$，则 $y_i \sim N_p(\mu^*, \Sigma^*)$，其中 $\mu^* = D(\mu-\mu_0)$，$\Sigma^* = D\Sigma D'$。此时(5.9)转化为

$$
\begin{aligned}
&H_0: \mu^*=0,\ \Sigma^*=I,\\
&H_1: \mu^*=0, \Sigma^*=I \text{ 至少有一个等式不成立}
\end{aligned} \tag{5.10}
$$

类似于(一)，利用似然比方法便可证明如下结果

定理 5.5 设 $x_i, i=1,2,\cdots,n$，是取自总体 $N_p(\mu,\Sigma)$ 的样本 $(n>p)$，其中 μ 和 $\Sigma(\Sigma>0)$ 都未知，检验假设(5.9)的似然比统计量是

$$
\begin{aligned}
\lambda = \left(\frac{e}{n}\right)^{\frac{np}{2}} |\Sigma_0^{-1}A|^{\frac{n}{2}} \exp\Big\{-\frac{1}{2}\operatorname{tr}\Sigma_0^{-1}\\
\cdot[A+n(\bar{x}-\mu_0)(\bar{x}-\mu_0)']\Big\}
\end{aligned} \tag{5.11}
$$

从上面定理可发现，统计量(5.11)正是检验假设 $H_0: \Sigma=\Sigma_0$，$H_1: \Sigma \neq \Sigma_0$ 的似然比统计量

$$
\lambda_1 = \left(\frac{e}{n}\right)^{\frac{np}{2}} |\Sigma_0^{-1}A|^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}[\Sigma_0^{-1}A]\right\}
$$

与在 $\Sigma=\Sigma_0$ 已知时检验假设 $H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0, H_1:\boldsymbol{\mu}\neq\boldsymbol{\mu}_0$ 的检验统计量

$$\lambda_2=\exp\left\{-\frac{1}{2}n(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'\Sigma_0^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)\right\}$$

的乘积，即 $\lambda=\lambda_1\cdot\lambda_2$。注意到 A 与 $\bar{\boldsymbol{x}}$ 相互独立，而 λ_1 仅是 A 的函数，λ_2 仅是 $\bar{\boldsymbol{x}}$ 的函数，所以 λ_1 与 λ_2 是相互独立的。还可以证明以(5.11)为基础的检验是无偏的.

由于 $\lambda=\lambda_1\cdot\lambda_2$，且 λ_1 与 λ_2 独立，从而

$$-2\ln\lambda=-2\ln\lambda_1+n(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'\Sigma_0^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)$$

在 H_0 为真时，上式右端第二项服从 $\chi^2(p)$ 分布，第一项由定理 5.2 知渐近服从 $\chi^2\left(\frac{p(p+1)}{2}\right)$ 分布.从而有

定理 5.6 在(5.9)中原假设 $H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0,\ \Sigma=\Sigma_0$ 为真时，统计量 $-2\ln\lambda$ 渐近服从自由度为 $\frac{1}{2}p(p+1)+p=\frac{1}{2}p(p+3)$ 的 χ^2 分布.

（四） 多个协方差矩阵相等的检验

设 $\boldsymbol{x}_i^{(k)}, i=1,2,\cdots,n_k$ 是从第 k 个独立正态总体 $N_p(\boldsymbol{\mu}^{(k)},\Sigma_k)$ 抽取的容量为 $n_k(n_k>p)$ 的样本。其中 $\boldsymbol{\mu}^{(k)}$ 和 $\Sigma_k(\Sigma_k>0)$ $k=1,2,\cdots,q$ 都是未知的. 我们要检验的假设是

$$H_0:\Sigma_1=\Sigma_2=\cdots=\Sigma_q(=\Sigma),$$

$$H_1:\Sigma_k,k=1,2,\cdots,q \text{ 不全相等} \tag{5.12}$$

其中 Σ 是某一个未知的正定矩阵.

记

$$\bar{\boldsymbol{x}}^{(k)}=\frac{1}{n_k}\sum_{i=1}^{n_k}\boldsymbol{x}_i^{(k)}\quad A_k=\sum_{i=1}^{n_k}(\boldsymbol{x}_i^{(k)}-\bar{\boldsymbol{x}}^{(k)})(\boldsymbol{x}_i^{(k)}-\bar{\boldsymbol{x}}^{(k)})'$$

$$A=\sum_{k=1}^{q}A_k\quad n=\sum_{k=1}^{q}n_k$$

定理 5.7 设 $\boldsymbol{x}_i^{(k)}, i=1,2,\cdots,n_k$ 为来自总体 $N_p(\boldsymbol{\mu}^{(k)},\Sigma_k)$

的容量为 $n_k(n_k>p)$ 的样本，$k=1,2,\cdots,q$，且这 q 个样本相互独立，则检验假设(5.12)的似然比统计量为

$$\lambda=\frac{\prod_{k=1}^{q}|A_k|^{\frac{n_k}{2}}}{|A|^{\frac{n}{2}}}\cdot\frac{n^{\frac{np}{2}}}{\prod_{k=1}^{q}n_k^{\frac{n_kp}{2}}} \tag{5.13}$$

证明 此时参数空间

$\Theta=\{(\boldsymbol{\mu}^{(1)},\cdots,\boldsymbol{\mu}^{(q)};\Sigma_1,\cdots,\Sigma_q):\boldsymbol{\mu}^{(k)}$ 为任意 p 维向量，

$\Sigma_k>0,k=1,2,\cdots,q\}$

相应于 H_0 的参数子空间

$\Theta_0=\{(\boldsymbol{\mu}^{(1)},\cdots,\boldsymbol{\mu}^{(q)};\Sigma_1,\cdots,\Sigma_q);\boldsymbol{\mu}^{(k)}$ 为任意 p 维实向量，

$\Sigma_k=\Sigma>0,k=1,2,\cdots,q\}$

似然函数是

$$\begin{aligned}L(\boldsymbol{\mu}^{(1)},\cdots,\boldsymbol{\mu}^{(q)};\Sigma_1,\cdots,\Sigma_q)&=\prod_{k=1}^{q}L(\boldsymbol{\mu}^{(k)};\Sigma_k)\\&=(2\pi)^{-\frac{np}{2}}\prod_{k=1}^{q}|\Sigma_k|^{-\frac{n_k p}{2}}\exp\Big\{-\frac{1}{2}\\&\quad\cdot\operatorname{tr}\Big[\Sigma_k^{-1}\cdot\sum_{i=1}^{n_k}(\boldsymbol{x}_i^{(k)}-\boldsymbol{\mu}^{(k)})\cdot(\boldsymbol{x}_i^{(k)}-\boldsymbol{\mu}^{(k)})'\Big]\Big\}\end{aligned}$$

因为每个因子 $L(\boldsymbol{\mu}^{(k)},\Sigma_k)$ 只与 $\boldsymbol{\mu}^{(k)}$ 和Σ_k有关，故要使 $L(\boldsymbol{\mu}^{(1)},\cdots,\boldsymbol{\mu}^{(q)};\Sigma_1,\cdots,\Sigma_q)$ 极大化，只须对各个因子 $L(\boldsymbol{\mu}^{(k)},\Sigma_k)$ 极大化，即有

$$\begin{aligned}\max_{\Theta}L(\boldsymbol{\mu}^{(1)},\cdots,\boldsymbol{\mu}^{(q)};\Sigma_1,\cdots,\Sigma_q)&=\prod_{k=1}^{q}L\left(\bar{\boldsymbol{x}}^{(k)},\frac{A_k}{n_k}\right)\\&=\prod_{k=1}^{q}(2\pi)^{-\frac{n_kp}{2}}\cdot\left|\frac{A_k}{n_k}\right|^{-\frac{n_k}{2}}\exp\left\{-\frac{n_kp}{2}\right\}\\&=(2\pi)^{-\frac{np}{2}}\exp\left\{-\frac{np}{2}\right\}\cdot\prod_{k=1}^{q}\left|\frac{A_k}{n_k}\right|^{-\frac{n_k}{2}}\end{aligned}$$

$$\max_{\Theta_0} L(\boldsymbol{\mu}^{(1)},\cdots,\boldsymbol{\mu}^{(q)};\Sigma)=\prod_{k=1}^{q} L\left(\bar{\boldsymbol{x}}^{(k)},\frac{A}{n}\right)$$

$$=(2\pi)^{-\frac{np}{2}}\left|\frac{A}{n}\right|^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}np\right\}$$

于是

$$\lambda=\frac{\max\limits_{\Theta_0} L(\boldsymbol{\mu}^{(1)},\cdots,\boldsymbol{\mu}^{(q)};\Sigma)}{\max\limits_{\Theta} L(\boldsymbol{\mu}^{(1)},\cdots,\boldsymbol{\mu}^{(q)};\Sigma_1,\cdots,\Sigma_q)}=\frac{\left|\frac{A}{n}\right|^{-\frac{n}{2}}}{\prod\limits_{k=1}^{q}\left|\frac{A_k}{n_k}\right|^{-\frac{n_k}{2}}}$$

$$=\frac{\prod\limits_{k=1}^{q}|A_k|^{\frac{n_k}{2}}}{|A|^{\frac{n}{2}}}\cdot\frac{n^{\frac{np}{2}}}{\prod\limits_{k=1}^{q}n_k^{\frac{n_kp}{2}}}$$

以 λ 为基础的似然比检验不是无偏的，但如果将 λ 中的 n_k 修改为 n_k-1，从而 n 变成 $n-q$，经过修改后的 λ 记为 λ^*，即

$$\lambda^*=\frac{\prod\limits_{k=1}^{q}|A_k|^{\frac{n_k-1}{2}}}{|A|^{\frac{n-q}{2}}}\cdot\frac{(n-q)^{\frac{(n-q)p}{2}}}{\prod\limits_{k=1}^{q}(n_k-1)^{\frac{(n_k-1)p}{2}}}\tag{5.14}$$

以 λ^* 为基础的检验是无偏的。下面是 Box 给出 λ^* 的有关渐近分布。

定理 5.8 在(5.12)中原假设 H_0 为真时，$-2\beta\ln\lambda^*$ 当 $n\to\infty$ 以及 $\lim\limits_{n\to\infty}\frac{n_k}{n}>0$ 时渐近地服从自由度为 $f=\frac{1}{2}p(p+1)(q-1)$ 的 χ^2 分布，其中

$$\beta=\begin{cases}1-\dfrac{2p^2+3p-1}{6(p+1)(q-1)}\left[\sum\limits_{k=1}^{q}\dfrac{1}{n_k-1}-\dfrac{1}{n-q}\right] & \text{当 } n_k \text{ 不全相等}\\[2ex] 1-\dfrac{(2p^2+3p-1)(q+1)}{6(p+1)(n-q)} & \text{当 } n_1=n_2=\cdots=n_q\end{cases}$$

在利用 T^2 统计量检验两个正态总体是否有相同的均值向量时，必须先确定这两个总体是否有相同的协方差阵，这时就要利用上述的检验。

例 5.2　三种不同的鸢尾花的两个指标：花瓣宽，花萼宽分别服从二元正态分布 $N_2(\boldsymbol{\mu}^{(i)},\Sigma_i)$ $(i=1,2,3)$，现分别抽取了容量为 $n_1=n_2=n_3=50$ 的样本，并计算得样本协方差矩阵分别为

$$S_1=\frac{1}{49}A_1=\begin{pmatrix}0.0661 & 0.0298\\ 0.0298 & 0.0601\end{pmatrix}$$

$$S_2=\frac{1}{49}A_2=\begin{pmatrix}0.0711 & 0.0333\\ 0.0333 & 0.0400\end{pmatrix}$$

$$S_3=\frac{1}{49}A_3=\begin{pmatrix}0.0588 & 0.0323\\ 0.0323 & 0.0810\end{pmatrix}$$

为了建立 Bayes 判别函数(参见第七章例 2.2)，首先必须检验 M_0: $\Sigma_1=\Sigma_2=\Sigma_3$ 是否成立。为此计算

$$|A_1|=49^2\begin{vmatrix}0.0661 & 0.0298\\ 0.0298 & 0.0601\end{vmatrix}=7.4061$$

$$|A_2|=49^2\begin{vmatrix}0.0711 & 0.0333\\ 0.0333 & 0.0400\end{vmatrix}=4.1660$$

$$|A_3|=49^2\begin{vmatrix}0.0588 & 0.0323\\ 0.0323 & 0.0810\end{vmatrix}=8.9931$$

$$|A|=|A_1+A_2+A_3|=49^2\begin{vmatrix}0.1960 & 0.0954\\ 0.0954 & 0.1811\end{vmatrix}=63.3731$$

此例中 $n=n_1+n_2+n_3=150$，$p=2$，$q=3$。并将它们代入(5.14)式计算得

$$\lambda^*=3.02856\times10^{-3}$$

而

$$\beta=1-\frac{(2p^2+3p-1)(q+1)}{6(p+1)(n-q)}=0.9804$$

所以

$$-2\beta\ln\lambda^* = 11.3673$$

查自由度 $f = \frac{1}{2}p(p+1)(q-1) = 6$。显著水平 $\alpha = 0.05$ 的 χ^2 分布表得临界值 $\chi^2_{0.95}(6) = 12.59$，显然

$$-2\beta\ln\lambda^* < \chi^2_{0.95}(6)$$

于是不能拒绝 H_0，即三个二维正态总体的协方差矩阵没有显著差异。

（五） 多个正态总体分布相同的检验

设 $x_i^{(k)}$，$i = 1,2,\cdots,n_k$，为来自第 k 个总体 $N_p(\mu^{(k)},\Sigma_k)$ 的样本 $k = 1,2,\cdots,q$，要检验的假设是

$$H_0: \mu^{(1)} = \mu^{(2)} = \cdots = \mu^{(q)} (= \mu),$$
$$\Sigma_1 = \Sigma_2 = \cdots = \Sigma_q (= \Sigma)$$
$$H_1: \mu^{(k)}, k = 1,2,\cdots,q \text{ 及 } \Sigma_k, k = 1,2,\cdots,q$$

至少有一个不全相等 (5.15)

记

$$\bar{x}^{(k)} = \frac{1}{n_k}\sum_{i=1}^{n_k} x_i^{(k)} \quad \bar{x} = \frac{1}{n}\sum_{k=1}^{q}\sum_{i=1}^{n_k} x_i^{(k)} \quad n = \sum_{k=1}^{q} n_k$$

$$A_k = \sum_{i=1}^{n_k}(x_i^{(k)} - \bar{x}^{(k)})(x_i^{(k)} - \bar{x}^{(k)})' \quad A = \sum_{k=1}^{q} A_k$$

$$W = \sum_{k=1}^{q}\sum_{i=1}^{n_k}(x_i^{(k)} - \bar{x})(x_i^{(k)} - \bar{x})'$$

$$= A + \sum_{k=1}^{q} n_k(\bar{x}^{(k)} - \bar{x})(\bar{x}^{(k)} - \bar{x})'$$

定理 5.9 检验假设(5.15)的似然比统计量为

$$\lambda = \frac{\prod_{k=1}^{q} |A_k|^{\frac{n_k}{2}}}{|W|^{\frac{n}{2}}} \cdot \frac{n^{\frac{np}{2}}}{\prod_{k=1}^{q} n_k^{\frac{n_k p}{2}}} \tag{5.16}$$

证明类似于定理 5.7. 读者可自己完成.

与(三)的情况相类似，若记检验假设 (5.12) 的似然比统计量 (5.13)为 λ_1，记当 Σ 相同时检验多个总体的均值是否相等(见第六章第 3 节多元方差分析)的检验统计量为 $\lambda_2 = |A|^{\frac{n}{2}}/|W|^{\frac{n}{2}}$，此时可以发现 $\lambda = \lambda_1 \cdot \lambda_2$. 在实际应用时，我们是采用(5.16)的修改形式，即在(5.16)式中用 $n_k - 1$ 代替 n_k，$n - q$ 代替 n. 修改后的统计量记为

$$\lambda^* = \frac{\prod_{k=1}^{q} |A_k|^{\frac{n_k-1}{2}} (n-q)^{\frac{(n-q)p}{2}}}{|W|^{\frac{n-q}{2}} \prod_{k=1}^{q} (n_k - 1)^{\frac{(n_k-1)p}{2}}} \tag{5.17}$$

定理 5.10 在 (5.15) 中 H_0 为真时，$-2\beta \ln \lambda^*$ 当 $n \to \infty$ 及 $\lim_{n\to\infty} \frac{n_k}{n} > 0$ 时渐近地服从自由度 $f = \frac{1}{2}(q-1)p(p+3)$ 的 χ^2 分布，其中 λ^* 为(5.17)所示，而

$$\beta = 1 - \left(\sum_{k=1}^{q} \frac{1}{n_k - 1} - \frac{1}{n-q}\right) \cdot \left(\frac{2p^2 + 3p - 1}{6(q-1)(p+3)}\right) + \frac{p - q + 2}{(n-q)(p+3)}$$

表 5.1 甲、乙两地段样地的三个树种的蓄积量数据

	样地号	1	2	3	4	5	6	7	8	9	10	11	12
甲地段	针叶乔木 x_1	3.6	1.7	2.1	0.5	1.3	0	5.0	1.5	6.4	4.1	1.9	1.9
	阔叶乔木 x_2	8.9	4.7	5.0	5.4	9.0	7.4	6.8	10.1	6.6	6.5	4.7	5.7
	亚乔木 x_3	0.7	1.0	2.0	2.0	1.0	4.3	0.2	1.1	0.3	0.3	3.4	0.7
	样地号	1	2	3	4	5	6	7	8	9	10	11	12
乙地段	针叶乔木 x_1	2.1	1.4	2.8	1.9	6.7	1.7	0	3.6	2.7	4.0	2.3	1.3
	阔叶乔木 x_2	8.6	8.9	6.6	8.1	10.1	5.1	4.4	5.3	5.2	8.1	6.8	5.9
	亚乔木 x_3	0.3	1.0	1.0	1.3	0	1.7	4.2	0.1	1.4	0.5	2.3	4.7

例 5.3 为了鉴别两地段上林分结构是否相同，在每个地段设置了 12 个样地(即每个地段抽取了容量为 12 的样本)分别对针叶乔木、阔叶乔木及亚乔木测定了样地内的蓄积量，数据列于表 5.1。所谓林分结构是否相同就是要比较树种组的组成是否相同，即要检验二地段树种蓄积量的平均是否相同；同时还要比较树种的相关系是否相同，即要检验二地段树种蓄积量的协方差阵是否相同。于是我们的原假设为 $H_0:\boldsymbol{\mu}^{(1)}=\boldsymbol{\mu}^{(2)},\ \Sigma_1=\Sigma_2$.

由表 5.1 的数据可计算出

$$A_1=\sum_{i=1}^{12}(\boldsymbol{x}_i^{(1)}-\bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}_i^{(1)}-\bar{\boldsymbol{x}}^{(1)})'$$

$$=\begin{pmatrix} 39.44 & 0.73 & -18.19 \\ 0.73 & 35.81 & -5.55 \\ -18.19 & -5.55 & 18.38 \end{pmatrix}$$

$$A_2=\sum_{i=1}^{12}(\boldsymbol{x}_i^{(2)}-\bar{\boldsymbol{x}}^{(2)})(\boldsymbol{x}_i^{(2)}-\bar{\boldsymbol{x}}^{(2)})'$$

$$=\begin{pmatrix} 31.31 & 18.35 & -19.29 \\ 18.35 & 36.04 & -17.32 \\ -19.29 & -17.32 & 25.39 \end{pmatrix}$$

$$W=\sum_{k=1}^{2}\sum_{i=1}^{12}(\boldsymbol{x}_i^{(k)}-\bar{\boldsymbol{x}})(\boldsymbol{x}_i^{(k)}-\bar{\boldsymbol{x}})'$$

$$=\begin{pmatrix} 70.70 & 19.13 & -37.45 \\ 19.13 & 72.07 & -22.53 \\ -37.45 & -22.53 & 43.90 \end{pmatrix}$$

再计算它们对应行列式的值：$|A_1|=13033.00$，$|A_2|=9633.57$，$|W|=102970.42$。 $\beta=1-0.00205=0.99795$，$q=2$，$n_1=n_2=12$。代入统计量 $-2\beta\ln\lambda^*$ 得

$$-2\beta\ln\lambda^*=3.04$$

对于 $\alpha=0.05$，查自由度 $f=\frac{1}{2}(2-1)\cdot3\cdot(3+1)=6$ 的 χ^2

分布表得 $\chi^2_{0.95}(6) = 12.592$，显然 $-2\beta\ln\lambda^* < \chi^2_{0.95}(6)$，所以不能拒绝 H_0，即认为甲乙两地段林分结构相同。

（六） **独立性检验**

一个 p 维随机向量划分为两个子向量，即 $\boldsymbol{x} = (\boldsymbol{x}^{(1)\prime}, \boldsymbol{x}^{(2)\prime})'$，如果子向量 $\boldsymbol{x}^{(1)}$ 与 $\boldsymbol{x}^{(2)}$ 独立，则在处理多元统计分析中的许多问题会带来极大的方便。 对于正态而言，$\boldsymbol{x}^{(1)}$ 与 $\boldsymbol{x}^{(2)}$ 的独立等价于它们的协方差矩阵 $\mathrm{cov}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = 0$。 因此检验 $\boldsymbol{x}^{(1)}$ 与 $\boldsymbol{x}^{(2)}$ 的独立性等价于检验其协方差阵等于零矩阵。下面就更一般的问题进行讨论。

设 $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$，$\Sigma > 0$，将 $\boldsymbol{x}$ 划分成 q 个子向量，其均值向量 $\boldsymbol{\mu}$ 和协方差阵 Σ 也进行相应划分：

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}^{(1)} \\ \vdots \\ \boldsymbol{x}^{(q)} \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \vdots \\ \boldsymbol{\mu}^{(q)} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1q} \\ \vdots & & \vdots \\ \Sigma_{q1} & \cdots & \Sigma_{qq} \end{pmatrix}$$

其中 $\boldsymbol{x}^{(k)}$ 的维数为 p_k，$k = 1, 2, \cdots, q$，且 $\sum_{k=1}^{q} p_k = p$。 我们要考查子向量 $\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(q)}$ 之间的相互独立性，它等价于检验假设

$$H_0: \Sigma_{kj} = 0 \quad k \neq j, \quad H_1: \ \Sigma_{kj} \text{ 不全为 0 矩阵}, \ k \neq j \qquad (5.18)$$

设 $\boldsymbol{x}_i$，$i = 1, 2, \cdots, n$ 为 $N_p(\boldsymbol{\mu}, \Sigma)$ 的容量为 $n(n > p)$ 的样本，$\boldsymbol{\mu}$ 与 $\Sigma(\Sigma > 0)$ 都未知。对于 $\boldsymbol{x}_i(i = 1, 2, \cdots, n)$，样本均值向量 $\bar{\boldsymbol{x}}$ 及 $A = \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$ 作相应划分：

$$\boldsymbol{x}_i = \begin{pmatrix} \boldsymbol{x}_i^{(1)} \\ \vdots \\ \boldsymbol{x}_i^{(q)} \end{pmatrix} \ (i = 1, 2, \cdots, n)$$

$$\bar{\boldsymbol{x}} = \begin{pmatrix} \bar{\boldsymbol{x}}^{(1)} \\ \vdots \\ \bar{\boldsymbol{x}}^{(q)} \end{pmatrix} \quad A = \begin{pmatrix} A_{11} & \cdots & A_{1q} \\ \vdots & & \\ A_{q1} & \cdots & A_{qq} \end{pmatrix}$$

利用多次应用的似然比方法可证得下述定理。

定理 5.11 设 $x_i(i=1,2,\cdots,n)$ 为正态总体 $N_p(\mu,\Sigma)$ 的样本 $(n>p)$，其中 μ 和 $\Sigma(\Sigma>0)$ 都是未知，检验假设 (5.18) 的似然比统计量为

$$\lambda^* = \left(\frac{|A|}{\prod_{k=1}^{q}|A_{kk}|}\right)^{\frac{n}{2}} \tag{5.19}$$

通常采用的是等价统计量

$$\lambda = \lambda^{*\frac{2}{n}} = \frac{|A|}{\prod_{k=1}^{q}|A_{kk}|} \tag{5.20}$$

λ 也可以用样本相关系数 R 来表示，将 R 也作相应的划分。记 a_{ij} 和 r_{ij} 分别为 A 和 R 的 (i,j) 元素，容易看出 $r_{ij}=\dfrac{a_{ij}}{\sqrt{a_{ii}\cdot a_{jj}}}$，因此有

$$|A| = |R|\cdot\prod_{i=1}^{p}a_{ii} \quad |A_{kk}| = |R_{kk}|\prod_{j\in\Omega}a_{jj}$$

其中

$$\Omega = \{(p_1+p_2+\cdots+p_{k-1}+1),(p_1+p_2+\cdots+p_{k-1}+2),\cdots,(p_1+p_2+\cdots+p_{k-1}+p_k)\}$$

将它们代入(5.20)便得到

$$\lambda = \frac{|R|}{\prod_{k=1}^{q}|R_{kk}|} \tag{5.21}$$

在 H_0 为真时 λ 的分布是什么? Anderson 的书[7]指出 λ 的分布与一些相互独立的 β 变量的乘积的分布相同，Box 1949 年证明了下述结论。

定理 5.12 在 (5.18) 中 H_0 为真时，$-m\ln\lambda$ 当 $n\to\infty$ 时渐近服从自由度为 $f=\dfrac{1}{2}\left(p^2-\sum_{k=1}^{q}p_k^2\right)$ 的 χ^2 分布，其中

$$m = n - 3 - \frac{p^3 - \sum_{k=1}^{q} p_k^3}{3\left(p^2 - \sum_{k=1}^{q} p_k^2\right)}$$

特别是在 $q = 2, p_1 = 1, p_2 = p - 1$ 时，由第二章(3.8)式知道，总体的第一个分量 x_1 与其余分量 $(x_2, \cdots, x_p)'$ 的复相关系数为 $\tilde{\rho} = \sqrt{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}/\sigma_{11}}$，所以独立性的假设 $H_0: \Sigma_{12} = 0$ 等价于 $H_0: \tilde{\rho} = 0$。即在此情况下问题化为对总体的一个分量与其它分量的复相关系数等于0的检验。由第四章定理2.7知，在 H_0 为真时有

$$\frac{n-p}{p-1} \cdot \frac{\tilde{r}^2}{1-\tilde{r}^2} \sim F(p-1, n-p)$$

其中 $\tilde{r}^2 = A_{12}A_{22}^{-1}A_{21}/a_{11}$，$A_{12} = (a_{12}, \cdots, a_{1p})$。于是，对于显著水平 α，由 F 分布表可查得临界值 $F_{1-\alpha}(p-1, n-p)$ 根据

$$\tilde{r}^2 \geqslant \left(1 + \frac{n-p}{(p-1)F_{1-\alpha}(p-1, n-p)}\right)^{-1}$$

成立与否来决定拒绝还是接受原假设 $H_0: \tilde{\rho} = 0$。

如果更特殊地当 $q = 2, p_1 = 1, p_2 = 1$，这时独立性的检验正是二元正态总体相关系数 $\rho = 0$ 的检验问题。类似地可利用第四章定理2.4，当 $H_0: \rho = 0$ 为真时有

$$t = \frac{\sqrt{n-2}\,r}{\sqrt{1-r^2}} \sim t(n-2)$$

其中 $r = a_{12}/\sqrt{a_{11}a_{22}}$，由给定的显著水平 α 及自由度 $n-2$ 查 t 分布表得临界值 $t_{1-\frac{\alpha}{2}}(n-2)$ 根据 $|t| \geqslant t_{1-\frac{\alpha}{2}}(n-2)$ 成立与否决定拒绝还是接受原假设。

与此类似地也可以对偏相关系数是否为0作出相应的假设检验，此时只须利用第四章推论2.2作假设检验。

例5.4　对河水进行了12次测量，测量的对象是 x_1：水中氧气含量的自然对数，x_2：水流速度的自然对数，x_3：水深度的自然对

数. 由所得数据计算其样本协方差阵和相关系数阵分别为:

$$S=\begin{pmatrix}7.250 & 0.507 & -3.878\\ & 0.480 & -0.056\\ & & 2.550\end{pmatrix},\ R=\begin{pmatrix}1 & 0.272 & -0.902\\ & 1 & -0.051\\ & & 1\end{pmatrix}$$

首先,我们检验 x_1 与 x_2 的相关系数 ρ_{12} 是否等于 0. 计算

$$t=\sqrt{n-2}\cdot r_{12}/\sqrt{1-r_{12}^2}$$
$$=\sqrt{10}\times 0.272/\sqrt{1-0.272^2}=0.8938$$

对于 $\alpha=0.05$ 查自由度为 10 的 t 分布表得临界值 $t_{1-\frac{\alpha}{2}}(10)=t_{0.975}(10)=2.23$,显然 $|t|=0.8938<2.23$,于是 $\rho_{12}=0$ 不能拒绝.

其次,我们检验在 x_3 给定时 x_1 与 x_2 的偏相关系数 $\rho_{12\cdot 3}$ 是否等于 0. 先计算偏协方差矩阵

$$S_{1\cdot 2}=S_{11}-S_{12}S_{22}^{-1}S_{21}$$
$$=\begin{pmatrix}7.25 & 0.507\\ 0.507 & 0.480\end{pmatrix}-\begin{pmatrix}-3.878\\ -0.056\end{pmatrix}$$
$$\cdot(2.55)^{-1}(-3.878\quad -0.056)$$
$$=\begin{pmatrix}1.35 & 0.42\\ 0.42 & 0.48\end{pmatrix}$$

由此可计算出样本偏相关系数 $r_{12\cdot 3}=0.42/\sqrt{1.35\times 0.48}=0.52$

根据第四章推论 2.2,知道在 $\rho_{12\cdot 3}=0$ 为真时,

$$t_1=\sqrt{n-3}\,r_{12\cdot 3}/\sqrt{1-r_{12\cdot 3}^2}\sim t(n-3)$$

将 $r_{12\cdot 3}=0.52$ 代入上述 t_1 统计量得 $t_1=1.826$, 对于 $\alpha=0.05$ $t_{0.975}(9)=2.26$,显然 $|t_1|<t_{0.975}(9)$,所以不能拒绝 $\rho_{12.3}=0$.

最后,检验 x_1 与 (x_2x_3) 的复相关系数 $\bar{\rho}=\rho_{1\cdot 23}$ 是否等于 0. 此时可计算样本复相关系数 $\bar{r}^2=\frac{S_{12}S_{22}^{-1}S_{21}}{S_{11}}=0.864$, 并代入统计量 $F=\frac{n-p}{p-1}\cdot\frac{\bar{r}^2}{1-\bar{r}^2}=28.6$.对于 $\alpha=0.05$,临界值为 $F_{0.95}(2,$

9) = 4.26，显然 $F > F_{0.95}(2,9)$ 所以 $\bar{\rho} = 0$ 的假设应该拒绝.

上面讨论说明了：水速对氧气含量没有显著影响，即使当水深给定后也是这样. 由于 $S_{11} = S_{1\cdot 2} + \bar{r}^2 S_{11}$，即水速和水深影响了氧含量的方差的 86.4%，而由于 $r_{13}^2 = 0.814$，水深影响了氧含量的方差的 81.4%. 由于 $\bar{\rho} = 0$ 的假设应拒绝，从而氧气含量与水速、水深之间显著存在着线性统计关系.

在本节给出了许多有关协方差阵的似然比检验统计量的渐近分布. 读者可能已注意到，它们都是形如 $-\rho\ln\lambda^* \sim \chi^2(f)$ 的形式，这里的 ρ 对于不同的似然比统计量是不同的，然而 χ^2 分布的自由度 f 是有一定规律的，它等于在 Θ 下未知参数的个数与在 Θ_0 下未知参数的个数之差. 例如在(一)中检验假设 $H_0: \Sigma = \Sigma_0; H_1: \Sigma \neq \Sigma_0$. 在 Θ 中 Σ 的未知参数有 $\frac{1}{2}p(p+1)$，μ 的未知参数个数为 p，总共有 $\frac{1}{2}p(p+1)+p$ 个未知参数，而在 Θ_0 中未知参数仅有 μ 的 p 个未知参数. 所以自由度 $f = \frac{1}{2}p(p+1)$，这与定理 5.2 是一致的. 其它情况也是如此，读者可以自己去验证. 上述规律对第六章中回归系数矩阵的检验也同样成立.

习　题　五

5.1　为了检验某化验员的化学分析是否有系统误差，今取四个等级的铁矿石标样. 四个等级的铁矿石的铁的含量用四维向量 $x = (x_1, x_2, x_3, x_4)'$ 表示. 设其均值向量为 $Ex = \mu_0 = (22.75, 32.75, 51.50, 61.50)$，并假定 x 为四维正态向量. 让化验员对标样进行分析，每次化验从低到高，重复化验 21 次，根据化验的数据计算得

$$\bar{x} = (22.82,\ 32.79,\ 51.45,\ 61.38)'$$

$$A = (n-1)S = \begin{pmatrix} 0.702 & & & \\ 0.541 & 0.712 & & \\ 0.184 & 0.228 & 0.392 & \\ 0.253 & 0.258 & 0.346 & 0.806 \end{pmatrix}$$

(a) 对于 $\alpha = 0.05$，检验化验员的分析与 μ_0 是否有显著差异?

(b) 若有显著差异，分别用一元的 t 检验法对四个均值分量分别检验，以帮助判定(a)中的差异是来自哪一个等级标样 ($\alpha = 0.05$)。

(c) 求出化验员在四个等级中化验均值的 0.95 Bonferroni 置信区间 (取 $t_{1-\frac{0.05}{8}}(20) = 2.607$)。

5.2 （成对比较）有时两总体的样品必须成对地出现才能作比较。设 $\{x_i, y_i\}$，$i = 1, 2, \cdots, n$ 是成对的试验数据，令 $z_i = x_i - y_i$，$i = 1, 2, \cdots, n$，且独立，并服从 $N_p(\delta, \Sigma)$($\Sigma > 0$，未知)。

(a) 用似然比方法导出检验假设 $H_0: \delta = 0, H_1: \delta \neq 0$ 的检验统计量。

(b) 导出 δ 的 p 个分量的 Bonferroni 置信区间。

5.3 依照环境保护法，污水必须经过处理达到规定标准才能排放。某工厂化验室及环保部门化验室对该厂的 11 个水样进行测试，测试两个指标：x_1 = 生化氧，x_2 = 悬浮固体量。其数据如下表：

化验室 \ 样号		1	2	3	4	5	6	7	8	9	10	11
厂化验室	x_1	6	6	18	8	11	34	28	71	43	33	20
	x_2	27	23	64	44	30	75	26	124	54	30	14
环保化验室	x_1	25	28	36	35	15	44	42	54	34	29	39
	x_2	15	13	22	29	31	64	30	64	56	20	21

(a) 利用习题 5.2 给出的成对比较法检验两家化验室的分析结果是否一致? ($\alpha = 0.05$)

(b) 若不一致，求出 $\delta = [\delta_1, \delta_2]' = [\mu_{11} - \mu_{21}, \mu_{12} - \mu_{22}]'$ 的 0.95 Bonferroni 置信区间，其中 μ_{i1}, μ_{i2}，$i = 1, 2$ 分别表示两个化验室所测的生化氧及悬浮固体量的总体均值。

5.4 设 $x \sim N_p(\mu, \Sigma)$，$\Sigma > 0$ 且未知 x_i $i = 1, 2, \cdots, n$ 是从该总体抽取的容量为 $n(n > p)$ 的样本，C 为已知的 $k \times p$ 阶矩阵，$k < p$，$\mathrm{rk}(C) = k$，μ_0 为已给定的 p 维向量，要检验假设

$$H_0: C\mu = C\mu_0, \quad H_1: C\mu \neq C\mu_0$$

试导出这个检验的似然比统计量。

5.5 为了判断两个不同产地的鸢尾花是否属于一个种，各取了 50 个样品测量了 x_1 = 花的萼片长，x_2 = 萼片宽，x_3 = 花瓣长，x_4 = 花瓣宽，希望比较这些形态上的差异是否显著，作为一个分类的依据，由测得的数据算得两地的样本均值向量为

$$\bar{x}^{(1)} = (5.936,\ 2.770,\ 4.260,\ 1.326)'$$
$$\bar{x}^{(2)} = (5.006,\ 3.428,\ 1.462,\ 0.246)'$$

及

$$A_1 = \sum_{i=1}^{50} (x_i^{(1)} - \bar{x}^{(1)})(x_i^{(1)} - \bar{x}^{(1)})' = \begin{pmatrix} 9.6012 & & & \\ 5.0067 & 5.8823 & & \\ 4.7654 & 2.8001 & 6.1919 & \\ 1.8011 & 1.3025 & 1.9026 & 1.3118 \end{pmatrix}$$

$$A_2 = \sum_{i=1}^{50} (x_i^{(2)} - \bar{x}^{(2)})(x_i^{(2)} - \bar{x}^{(2)})' = \begin{pmatrix} 9.5422 & & & \\ 4.9289 & 5.9835 & & \\ 4.9980 & 1.8231 & 6.1059 & \\ 1.4383 & 1.1721 & 1.9768 & 1.1486 \end{pmatrix}$$

(*a*) 设两地鸢尾花形态的协方差阵相等，对于 $\alpha = 0.05$，判断两地鸢尾花形态均值向量是否有显著差异？

(*b*) 若有显著差异，求出两总体均值向量差各分量的 95% Bonferroni 置信区间。

5.6 设 $x = (x_1, x_2, x_3, x_4)'$ 表示从北、东、南、西四个方向钻入软木树所得的软木量。记 $Ex = \mu = (\mu_1, \mu_2, \mu_3, \mu_4)'$，设 x 服从正态分布。现观测了 10 株树，并计算了样本均值向量及协方差阵为

$$\bar{x} = (0.96,\ 0.93,\ 0.99,\ 0.93)'$$

$$S = \frac{1}{9} A = \frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x})(x_i - \bar{x})'$$

$$= \frac{1}{9} \begin{pmatrix} 0.704 & & & \\ 0.392 & 0.721 & & \\ 0.486 & 0.483 & 0.609 & \\ 0.352 & 0.401 & 0.393 & 0.541 \end{pmatrix}$$

对于 $\alpha = 0.05$，检验 $\mu_1 = \mu_2 = \mu_3 = \mu_4$ 是否成立。

*5.7 设 x_i, $i = 1, 2, \cdots, n\ (n > 2p)$ 是从正态总体 $N_{2p}(\mu, \Sigma)$ $(\Sigma > 0$ 且未知)抽取的样本，设

$$\mu=\begin{pmatrix}\mu^{(1)}\\ \mu^{(2)}\end{pmatrix},\quad \Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$$

其中 $\mu^{(1)}$, $\mu^{(2)}$ 是 p 维向量 Σ_{11},Σ_{22} 是 $p\times p$ 阶矩阵，试作出检验假设

$$H_0:\mu^{(1)}=\mu^{(2)},\quad H_1:\mu^{(1)}\neq\mu^{(2)}$$

的检验方案。

5.8 为了研究某城市男性高中学生吸烟问题，随机抽取了 300 个高中男生，调查每天吸烟量结果如下表

吸烟量	不吸烟	少量(1—4 支)	中等(5—10 支)	严重(10 支以上)
人数	168	85	30	17

试对这四种情况的人的比例 p_i, $i=1,2,3,4\left(\sum_{i=1}^{4}p_i=1\right)$ 作出 95% 的同时置信区间。

5.9 证明定理 5.3、5.9 和 5.11。

5.10 利用习题 5.3 的数据，检验厂化验室(第一总体)与环保部门化验室(第二总体)的协方差阵是否相等 ($\alpha=0.05$)。

5.11 利用习题 5.5 的数据检验两地的鸢尾花形态的协方差阵是否相等 ($\alpha=0.05$)。

5.12 设 x_i, $i=1,2,\cdots,n(n>p)$ 为总体 $N_p(\mu,\Sigma)$，($\Sigma>0,\mu,\Sigma$ 都未知)的样本，试导出检验假设

$$H_0:\Sigma=\mathrm{diag}(\sigma_{11},\sigma_{22},\cdots,\sigma_{pp}),\quad H_1:\Sigma\neq\mathrm{diag}(\sigma_{11},\sigma_{22},\cdots\sigma_{pp})$$

的似然比统计量为 $\lambda=|R|^{n/2}$，其中 R 为样本相关系数矩阵。

第六章　多元线性统计模型

在一元统计分析中，回归分析与方差分析是应用最广泛的统计方法．在多元统计分析中也是如此，多元回归分析与多元方差分析都属于多元线性统计模型．在本章中将介绍多个自变量与多个因变量的回归问题，它包括回归模型的参数估计、检验、回归预测以及自变量的筛选，最后再介绍多元方差分析．

§6.1　多元线性回归模型

在一元统计中也有多元回归分析之称，它指的是多个自变量对一个因变量(简称一对多)．设 x 为因变量，$z_1, z_2, \cdots, z_q$ 为 q 个自变量，一元统计中的线性回归模型为：

$$\begin{cases} x = \theta_0 + \theta_1 z_1 + \cdots + \theta_q z_q + \varepsilon \\ E(\varepsilon) = 0 \quad \operatorname{var}(\varepsilon) = \sigma^2. \end{cases} \tag{1.1}$$

若我们已得到 n 组独立观察值 $(x_i; z_{1i}, z_{2i}, \cdots, z_{qi})$，$i = 1, \cdots, n$．对于这 n 组观察值，模型(1.1)可写为

$$\begin{cases} x_i = \theta_0 + \theta_1 z_{1i} + \cdots + \theta_q z_{qi} + \varepsilon_i \\ E(\varepsilon_i) = 0, \ \operatorname{var}(\varepsilon_i) = \sigma^2 \\ i = 1, 2, \cdots, n \end{cases} \tag{1.2}$$

记

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \ \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_q \end{pmatrix}, \ \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \ Z = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_{11} & z_{12} & \cdots & z_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ z_{q1} & z_{q2} & \cdots & z_{qn} \end{pmatrix}$$

则模型(1.2)可写为

$$\begin{cases} x = Z'\theta + \epsilon \\ E(\epsilon) = o, \ \text{cov}(\epsilon) = \sigma^2 I_n \end{cases} \tag{1.3}$$

我们称 Z 为设计矩阵，且设 $\text{rk}Z = q + 1 < n$，称 x 为 n 维观察向量，θ 是 $q + 1$ 维未知参数，称为回归系数向量，ϵ 是 n 维非观察值，它代表随机误差，其分量 $\varepsilon_i(i = 1,2,\cdots,n)$ 相互独立同方差 σ^2(未知的)。众所周知，在一元统计中就是要在所给的 Z 及 x 之下对 θ 及 σ^2 进行估计、建立回归方程、讨论估计的有关性质，若在正态假设下(即假设 $x \sim N_n(Z'\theta, \sigma^2 I_n)$ 或 $\epsilon \sim N_p(o, \sigma^2 I_n)$)还要对回归系数进行检验及预测精度的讨论。

在多元统计中多元回归分析是指多个自变量对多个因变量(简称多对多)。在实际工作中常常要考察多对多的问题，例如在炼钢过程中，要考察钢的含碳量，含硫量，…等等与冶炼初期天然矿石品种，熔化时间，熔化温度，…等等之间的关系，这就是多对多问题；在地质学、气象学、生物学和医学中也都会碰到类似的多对多的问题。与一元统计中的一对多回归分析一样，首先给出多对多的线性回归模型。

设有 p 个因变量 $x_1, x_2, \cdots, x_p$；有 q 个自变量 $z_1, z_2, \cdots, z_q$。假定每个因变量 x_i $(i = 1,2,\cdots,p)$ 都有自己的线性回归模型，即

$$\begin{cases} x_1 = \theta_{01} + \theta_{11}z_1 + \cdots + \theta_{q1}z_q + \varepsilon_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_p = \theta_{0p} + \theta_{1p}z_1 + \cdots + \theta_{qp}z_q + \varepsilon_p \end{cases} \tag{1.4}$$

(1.4)式中的每一个等式就是一个一对多的回归模型，这里要注意的是(1.4)中每一个随机误差 $\varepsilon_i(i = 1,2,\cdots,p)$ 是相关的，即它们满足

$$E(\varepsilon_i) = 0, \ \text{cov}(\varepsilon_i, \varepsilon_j) = \sigma_{ij} \ (i,j = 1,2,\cdots,p) \tag{1.5}$$

为了今后讨论的方便可将(1.4)与(1.5)式写成矩阵形式：

$$\begin{cases} x = \Theta z + \epsilon & (1.6) \\ E(\epsilon) = o, \ \text{cov}(\epsilon) = \Sigma = (\sigma_{ij})_{p\times p} & (1.7) \end{cases}$$

其中

$$\boldsymbol{\Theta}=\begin{pmatrix}\theta_{01} & \theta_{11}\cdots\theta_{q1}\\ \theta_{02} & \theta_{12}\cdots\theta_{q2}\\ \vdots & \vdots \qquad \vdots\\ \theta_{0p} & \theta_{1p}\cdots\theta_{qp}\end{pmatrix}=(\boldsymbol{\theta}_0\ \boldsymbol{\theta}_1\cdots\boldsymbol{\theta}_q),$$

$$\boldsymbol{x}=\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_p\end{pmatrix},\ \boldsymbol{z}=\begin{pmatrix}1\\ z_1\\ \vdots\\ z_q\end{pmatrix},\ \boldsymbol{\varepsilon}=\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_p\end{pmatrix}.$$

如果对自变量 $\boldsymbol{z}$ 设计 n 组不同的值 $\boldsymbol{z}_j\ (j=1,2,\cdots,n)$，那么在 $\boldsymbol{z}_j$ 之下因变量 $\boldsymbol{x}$ 可得到 n 个观察向量 $\boldsymbol{x}_j(j=1,2,\cdots,n)$，且假设每次观察都是独立的．我们用矩阵形式表示．记

$$X=\begin{pmatrix}x_{11} & x_{12}\cdots x_{1n}\\ x_{21} & x_{22}\cdots x_{2n}\\ \cdots\cdots\cdots\cdots\\ x_{p1} & x_{p2}\cdots x_{pn}\end{pmatrix}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)=\begin{pmatrix}\boldsymbol{x}^{(1)'}\\ \boldsymbol{x}^{(2)'}\\ \vdots\\ \boldsymbol{x}^{(p)'}\end{pmatrix}\quad(1.8)$$

它是因变量观察矩阵，是 $p\times n$ 阶．

$$Z=\begin{pmatrix}1 & 1\cdots 1\\ z_{11} & z_{12}\cdots z_{1n}\\ \cdots\cdots\cdots\cdots\\ z_{q1} & z_{q2}\cdots z_{qn}\end{pmatrix}=(\boldsymbol{z}_1\ \boldsymbol{z}_2\cdots\boldsymbol{z}_n)=\begin{pmatrix}\boldsymbol{1}'\\ \boldsymbol{z}^{(1)'}\\ \vdots\\ \boldsymbol{z}^{(q)'}\end{pmatrix}\quad(1.9)$$

它是自变量设计矩阵，是 $(q+1)\times n$ 阶．

$$\Theta=\begin{pmatrix}\theta_{01} & \theta_{11}\cdots\theta_{q1}\\ \theta_{02} & \theta_{12}\cdots\theta_{q2}\\ \cdots\cdots\cdots\cdots\\ \theta_{0p} & \theta_{1p}\cdots\theta_{qp}\end{pmatrix}=(\boldsymbol{\theta}_0\ \boldsymbol{\theta}_1\cdots\boldsymbol{\theta}_q)=\begin{pmatrix}\boldsymbol{\theta}^{(1)'}\\ \boldsymbol{\theta}^{(2)'}\\ \vdots\\ \boldsymbol{\theta}^{(p)'}\end{pmatrix}\quad(1.10)$$

它是回归系数矩阵，是 $p\times(q+1)$ 阶．

$$\boldsymbol{\varepsilon}=\begin{pmatrix}\varepsilon_{11} & \varepsilon_{12}\cdots\varepsilon_{1n}\\ \varepsilon_{21} & \varepsilon_{22}\cdots\varepsilon_{2n}\\ \cdots\cdots\cdots\cdots\\ \varepsilon_{p1} & \varepsilon_{p2}\cdots\varepsilon_{pn}\end{pmatrix}=(\boldsymbol{\varepsilon}_1\ \boldsymbol{\varepsilon}_2\cdots\boldsymbol{\varepsilon}_n)=\begin{pmatrix}\boldsymbol{\varepsilon}^{(1)'}\\ \boldsymbol{\varepsilon}^{(2)'}\\ \vdots\\ \boldsymbol{\varepsilon}^{(p)'}\end{pmatrix}\quad(1.11)$$

它是误差矩阵，是 $p \times n$ 阶.

对于 n 次独立观察，模型(1.6)及(1.7)可写为

$$\begin{cases} X = \Theta Z + \boldsymbol{\epsilon} \\ \operatorname{cov}(\operatorname{vec}\boldsymbol{\epsilon}) = \Sigma \otimes I \text{ (或 } E(\boldsymbol{\epsilon}_j) = 0, \ \operatorname{cov}(\boldsymbol{\epsilon}_j) = \Sigma, \\ \qquad \operatorname{cov}(\boldsymbol{\epsilon}^{(i)}\boldsymbol{\epsilon}^{(j)}) = \sigma_{ij} I \text{)} \end{cases} \tag{1.12}$$

(1.12)式称为多对多线性回归模型. 在实际应用中往往称(1.8)及(1.9)式为资料矩阵. 如果当 $p = 1$ 时，则(1.12)变成

$$\boldsymbol{x}^{(1)'} = \boldsymbol{\theta}^{(1)'} Z + \boldsymbol{\epsilon}^{(1)'} \tag{1.13}$$

其中 $\boldsymbol{x}^{(1)'}, \boldsymbol{\theta}^{(1)'}, \boldsymbol{\epsilon}^{(1)'}$ 分别为 X, Θ 和 $\boldsymbol{\epsilon}$ 的第一行向量，经转置后就得到

$$\begin{cases} \boldsymbol{x}^{(1)} = Z' \boldsymbol{\theta}^{(1)} + \boldsymbol{\epsilon}^{(1)} \\ E(\boldsymbol{\epsilon}^{(1)}) = \boldsymbol{o}, \ \operatorname{cov}(\boldsymbol{\epsilon}^{(1)}) = \sigma_{11} I_n \end{cases} \tag{1.14}$$

其中 σ_{11} 是 Σ 的第一行第一列元素，即误差向量 $\boldsymbol{\epsilon}$ 的第一个分量的方差. (1.14)式正是一对多的回归模型(参看(1.3)式).

有了多对多回归模型(1.12)后，我们要做的工作是:

(1) 用最小二乘法估计回归系数矩阵 Θ，给出 Σ 的无偏估计，从而建立回归方程(预测方程)并讨论估计的性质.

(2) 在正态假设下，即在

$$X \sim N_{p,n}(\Theta Z, \Sigma, I_n) \tag{1.15}$$

之下给出 Θ 及 Σ 的极大似然估计，对回归系数进行检验以及讨论回归方程预测精度的估计.

(3) 讨论自变量的筛选问题.

多对多线性回归模型中 q 个自变量都是可以确定的实变量，这种模型用于建立回归方程，以便通过自变量的实现来对相应的因变量进行预测或控制. 如果模型中的自变量是二态变量(即它们只取 0 或 1)，它们的取值可由试验设计来确定，这种模型用于分析各个自变量对因变量的影响如何，这就是多元方差分析，我们将在本章的最后一节来介绍.

§6.2 多元线性回归模型的参数估计

(一) 最小二乘估计及其性质

由模型(1.12)可看出，对误差向量只要求二阶矩存在，无须对它的分布作出假设，在此情况下可利用最小二乘（简记 LS）方法估计 Θ。我们先回顾一下在一元统计中的 LS 估计，在一对多的模型 (1.3) 中设回归系数向量 $\boldsymbol{\theta}=(\theta_0,\theta_1,\cdots,\theta_q)'$ 的 LS 估计为 $\hat{\boldsymbol{\theta}}=(\hat{\theta}_0,\cdots,\hat{\theta}_q)'$，$\hat{\boldsymbol{\theta}}$ 是使得误差平方和

$$S(\boldsymbol{\theta})=(\boldsymbol{x}-Z'\boldsymbol{\theta})'(\boldsymbol{x}-Z'\boldsymbol{\theta})$$

达到极小，即 $\hat{\boldsymbol{\theta}}$ 满足

$$\begin{aligned}Q=(\boldsymbol{x}-Z'\hat{\boldsymbol{\theta}})'(\boldsymbol{x}-Z'\hat{\boldsymbol{\theta}})&=\min_{\boldsymbol{\theta}}(\boldsymbol{x}-Z'\boldsymbol{\theta})'(\boldsymbol{x}-Z'\boldsymbol{\theta})\\&=\min_{\boldsymbol{\theta}}S(\boldsymbol{\theta})\end{aligned}\tag{2.1}$$

上式中 Q 称为残差平方和，$(\boldsymbol{x}-Z'\hat{\boldsymbol{\theta}})$ 称为残差向量。由极值原理及第一章的(4.3)和(4.4)式有

$$\begin{aligned}\frac{\partial S(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}&=\frac{\partial}{\partial\boldsymbol{\theta}}(\boldsymbol{x}-Z'\boldsymbol{\theta})'(\boldsymbol{x}-Z'\boldsymbol{\theta})\\&=-2Z\boldsymbol{x}-2ZZ'\boldsymbol{\theta}=\boldsymbol{o}\end{aligned}$$

从而得到

$$ZZ'\boldsymbol{\theta}=Z\boldsymbol{x}\tag{2.2}$$

称(2.2)式为正规方程组，若 $|ZZ'|\neq 0$（即 $\operatorname{rk}Z=q+1$），则 $(ZZ')^{-1}$ 存在，这样便可求得 $\boldsymbol{\theta}$ 的 LS 估计为

$$\hat{\boldsymbol{\theta}}=(ZZ')^{-1}Z\boldsymbol{x}\tag{2.3}$$

或转置为

$$\hat{\boldsymbol{\theta}}'=\boldsymbol{x}'Z'(ZZ')^{-1}\tag{2.3$'$}$$

在多对多的模型(1.12)中对于每一个因变量都对应着一个一对多的模型

$$\begin{cases}\boldsymbol{x}^{(i)}=Z'\boldsymbol{\theta}^{(i)}+\boldsymbol{\varepsilon}^{(i)}\\E(\boldsymbol{\varepsilon}^{(i)})=\boldsymbol{o},\ \operatorname{cov}(\boldsymbol{\varepsilon}^{(i)})=\sigma_{ii}\boldsymbol{I}_n\\i=1,2,\cdots,p\end{cases}\tag{2.4}$$

其中 $\boldsymbol{x}^{(i)},\boldsymbol{\varepsilon}^{(i)},\boldsymbol{\theta}^{(i)}$ 由(1.8)~(1.11)式确定，σ_{ii} 为 Σ 的第 i 个对角线元素．对于(2.4)中每个 i 都有一个误差平方和

$$S_i(\boldsymbol{\theta}^{(i)})=(\boldsymbol{x}^{(i)}-Z'\boldsymbol{\theta}^{(i)})'(\boldsymbol{x}^{(i)}-Z'\boldsymbol{\theta}^{(i)})\quad(i=1,2,\cdots,p)$$

p 个误差平方和为

$$S(\Theta)=\sum_{i=1}^{p}S_i(\boldsymbol{\theta}^{(i)})=\sum_{i=1}^{p}(\boldsymbol{x}^{(i)}-Z'\boldsymbol{\theta}^{(i)})'(\boldsymbol{x}^{(i)}-Z'\boldsymbol{\theta}^{(i)})\tag{2.5}$$

多对多模型的回归系数矩阵 Θ 的 LS 估计 $\hat{\Theta}$ 就是使得(2.5)式达到极小．由于 $S_i(\boldsymbol{\theta}^{(i)})$ 都是平方和，因此要使 $S(\Theta)$ 达到极小等价于每个 $S_i(\boldsymbol{\theta}^{(i)})$ 达到极小，由(2.3)′知道对于每个 $S_i(\boldsymbol{\theta}^{(i)})$ 使其达到极小的 $\boldsymbol{\theta}^{(i)}$ 为

$$\hat{\boldsymbol{\theta}}^{(i)\prime}=\boldsymbol{x}^{(i)\prime}Z'(ZZ')^{-1}\quad(i=1,2,\cdots,p)\tag{2.6}$$

这样便得到 Θ 的 LS 估计为

$$\hat{\Theta}=\begin{pmatrix}\hat{\boldsymbol{\theta}}^{(1)\prime}\\ \hat{\boldsymbol{\theta}}^{(2)\prime}\\ \vdots\\ \hat{\boldsymbol{\theta}}^{(p)\prime}\end{pmatrix}=XZ'(ZZ')^{-1}\tag{2.7}$$

显然，当 $\mathrm{rk}Z=q+1$ 时，(2.7)式是方程

$$\Theta(ZZ')=XZ'\tag{2.8}$$

的唯一解，称(2.8)式为多对多的正规方程组．若 $\mathrm{rk}Z<q+1$ 时，正规方程组有不定解，可用广义逆表示如下：

$$\hat{\Theta}=XZ'(ZZ')^{-}\tag{2.9}$$

定理 2.1 模型(1.12)的回归系数矩阵 Θ 的 LS 估计 $\hat{\Theta}$ 满足

(1) 对于任何非零 p 维向量 $\boldsymbol{l}$ 有

$$\boldsymbol{l}'(X-\hat{\Theta}Z)(X-\hat{\Theta}Z)'\boldsymbol{l}=\min_{\Theta}\boldsymbol{l}'(X-\Theta Z)(X-\Theta Z)'\boldsymbol{l}\tag{2.10}$$

(2) 使误差的交叉乘积矩阵 $(X-\Theta Z)(X-\Theta Z)'$ 的迹达到最小，即

$$\mathrm{tr}(X-\hat{\Theta}Z)(X-\hat{\Theta}Z)'=\min_{\Theta}\mathrm{tr}(X-\Theta Z)(X-\Theta Z)'\tag{2.11}$$

(3) 使广义方差 $|(X-\Theta Z)(X-\Theta Z)'|$ 达到最小，即

$$|(X-\hat{\Theta}Z)(X-\hat{\Theta}Z)'| = \min_{\Theta}|(X-\Theta Z)(X-\Theta Z)'| \quad (2.12)$$

(4) 在矩阵非负定意义下，对任意 Θ，有

$$(X-\hat{\Theta}Z)(X-\hat{\Theta}Z)' \leqslant (X-\Theta Z)(X-\Theta Z)' \quad (2.13)$$

(5) 用 $\lambda_1(A)$ 表示矩阵 A 的最大特征根，对任意 Θ，有

$$\lambda_1((X-\hat{\Theta}Z)(X-\hat{\Theta}Z)') \leqslant \lambda_1((X-\Theta Z)(X-\Theta Z)') \quad (2.14)$$

证明　我们就(1)给出证明，(2)～(5)留给读者。

在(2.7)式两边右乘 (ZZ')，然后移项便得

$$XZ' - \hat{\Theta}(ZZ') = (X-\hat{\Theta}Z)Z' = 0 \quad (2.15)$$

从而也有

$$Z(X-\hat{\Theta}Z)' = 0 \quad (2.16)$$

对任意的 p 维非零向量 l，并注意到(2.15)及(2.16)式，有

$$\begin{aligned} & l'(X-\Theta Z)(X-\Theta Z)'l \\ = & l'[(X-\hat{\Theta}Z)+(\hat{\Theta}-\Theta)Z][(X-\hat{\Theta}Z)+(\hat{\Theta}-\Theta)Z]'l \\ = & l'(X-\hat{\Theta}Z)(X-\hat{\Theta}Z)'l \\ & + l'(\hat{\Theta}-\Theta)ZZ'(\hat{\Theta}-\Theta)'l \end{aligned} \quad (2.17)$$

注意到

$$l'(\hat{\Theta}-\Theta)ZZ'(\hat{\Theta}-\Theta)'l \geqslant 0$$

从而对一切 Θ 及(2.17)式，有

$$l'(X-\hat{\Theta}Z)(X-\hat{\Theta}Z)'l \leqslant l'(X-\Theta Z)(X-\Theta Z)'l$$

这样(1)便得证. ▌

根据上述定理，多对多回归模型(1.12)的回归系数矩阵 Θ 的 LS 估计是有多种含义的. 定理中的任一种情况都可作为 Θ 的 LS 估计的定义.

由正规方程组(2.8)求得 Θ 的 LS 估计 $\hat{\Theta}$ 后，这样便可建立多对多回归方程(预测方程)：

$$\hat{x} = \hat{\Theta}z \quad (2.18)$$

将(2.18)具体写出便是

$$\begin{cases} \hat{x}_1 = \hat{\theta}_{01} + \hat{\theta}_{11}z_1 + \cdots + \hat{\theta}_{q1}z_q \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{x}_p = \hat{\theta}_{0p} + \hat{\theta}_{1p}z_1 + \cdots + \hat{\theta}_{qp}z_q \end{cases} \quad (2.19)$$

将前面已知的设计矩Z的n个列向量 z_i $(i=1,2,\cdots,n)$ 分别代入(2.19)，并写成矩阵形式，就得到相应的回归值(预测值)矩阵

$$\hat{X}=\hat{\Theta}Z=XZ'(ZZ')^{-1}Z \tag{2.20}$$

实测值矩阵X与预测值矩阵$\hat{X}$之差

$$\hat{\varepsilon}=X-\hat{X}=X-XZ'(ZZ')^{-1}Z=X(I-Z'(ZZ')^{-1}Z) \tag{2.21}$$

称为残差矩阵。而称

$$\begin{aligned}Q&=\hat{\varepsilon}\hat{\varepsilon}'=(X-\hat{X})(X-\hat{X})'\\&=X(I-Z'(ZZ')^{-1}Z)(I-Z'(ZZ')^{-1}Z)X'\\&=X(I-Z'(ZZ')^{-1}Z)X'\end{aligned} \tag{2.22}$$

为残差交叉乘积矩阵。由于

$$(I-Z'(ZZ')^{-1}Z)Z'=Z'-Z'=0$$

再利用(2.20)、(2.21)式，有

$$\hat{\varepsilon}Z'=X(I-Z'(ZZ')^{-1}Z)Z'=0 \tag{2.23}$$

$$\hat{\varepsilon}\hat{X}'=X(I-Z'(ZZ')^{-1}Z)Z'(ZZ')^{-1}ZX'=0 \tag{2.24}$$

上面两式说明残差矩阵与设计矩阵、预测值矩阵的行向量之间是正交的，于是就有

$$XX'=\hat{X}\hat{X}'+\hat{\varepsilon}\hat{\varepsilon}' \tag{2.25}$$

注意到 $Z'(ZZ')^{-1}Z$ 正是在 $\mathscr{L}(Z')$ 上的投影阵 $P_{Z'}$，从而 $\hat{\varepsilon}$ 的每一行向量都是X的行向量在 $\mathscr{L}^{\perp}(Z')$ 上的投影，即

$$\hat{\varepsilon}'=P_{\mathscr{L}^{\perp}(Z')}X'$$

下面的定理 2.2 给出了 LS 估计 $\hat{\Theta}$ 及残差交叉乘积矩阵Q的有关性质。

定理 2.2 若 $\mathrm{rk}Z=q+1$，则

(1) $$E\hat{\Theta}=\Theta \tag{2.26}$$

$$EQ=(n-q-1)\Sigma \tag{2.27}$$

即$\hat{\Theta}$是Θ的无偏估计，$\frac{1}{n-q-1}Q$ 是Σ的无偏估计。

(2) $$\mathrm{cov}(\mathrm{vec}\hat{\Theta})=\Sigma\otimes(ZZ')^{-1} \tag{2.28}$$

即

$$\operatorname{cov}(\hat{\theta}^{(i)}\hat{\theta}^{(j)}) = \sigma_{ij}(ZZ')^{-1} \tag{2.29}$$

(3) $\hat{\Theta}$ 与 $\hat{\varepsilon}$ 互不相关.

证明 由模型(1.12),有

$$EX = E(\Theta Z + \varepsilon) = \Theta Z$$

所以

$$E\hat{\Theta} = E(XZ'(ZZ')^{-1}) = \Theta ZZ'(ZZ')^{-1} = \Theta$$

这便是(2.26)式,记

$$P = I - Z'(ZZ')^{-1}Z$$

显然 P 是投影阵,且 $\operatorname{rk} P = \operatorname{tr} P = n - q - 1$ 和 $ZP = PZ' = 0$. 记 Q 的元素为 q_{ij},由模型(1.12),有

$$E(\varepsilon) = E(X - \Theta Z) = 0, \text{ 且 } E(\varepsilon^{(i)}\varepsilon^{(j)'}) = \sigma_{ij}I_n$$

由(2.22)式可看出,Q 可表成

$$Q = XPX' = (X - \Theta Z)P(X - \Theta Z)' = \varepsilon P \varepsilon'$$

从而有

$$\begin{aligned} Eq_{ij} &= E(\varepsilon^{(j)'}P\varepsilon^{(i)}) = E(\operatorname{tr}\varepsilon^{(j)'}P\varepsilon^{(i)}) = E(\operatorname{tr}P\varepsilon^{(i)}\varepsilon^{(j)'}) \\ &= \operatorname{tr}[P \cdot E(\varepsilon^{(i)}\varepsilon^{(j)'})] = \operatorname{tr}[P\sigma_{ij}I_n] \\ &= \sigma_{ij}(n - q - 1) \end{aligned}$$

所以

$$EQ = (n - q - 1)\Sigma$$

这便证明了(2.27)式.

按第一章定理 1.1,由 $\hat{\Theta} = XZ'(ZZ')^{-1}$,所以

$$\operatorname{vec}\hat{\Theta} = (I \otimes (ZZ')^{-1}Z)\operatorname{vec} X$$

而由(1.12)知 $\operatorname{cov}(\operatorname{vec} X) = \Sigma \otimes I$,利用第二章定理 1.4,有

$$\begin{aligned} \operatorname{cov}(\operatorname{vec}\hat{\Theta}) &= (I \otimes (ZZ')^{-1}Z)\operatorname{cov}(\operatorname{vec} X)(I \otimes (ZZ')^{-1}Z)' \\ &= (I \otimes (ZZ')^{-1}Z)(\Sigma \otimes I)(I \otimes Z'(ZZ')^{-1}) \\ &= \Sigma \otimes (ZZ')^{-1} \end{aligned}$$

(2.28)式得证. 由于

$$\operatorname{vec}\hat{\varepsilon} = \operatorname{vec} XP = (I \otimes P)\operatorname{vec} X$$

所以

$$
\begin{aligned}
\operatorname{cov}(\operatorname{vec}\hat{\Theta},\operatorname{vec}\hat{\varepsilon}) &= \operatorname{cov}((I\otimes(ZZ')^{-1}Z),(I\otimes P)\operatorname{vec}X)\\
&= (I\otimes(ZZ')^{-1}Z)\cdot\operatorname{cov}(\operatorname{vec}X)\cdot(I\otimes P)\\
&= (I\otimes(ZZ')^{-1}Z)\cdot(\Sigma\otimes I)(I\otimes P)\\
&= \Sigma\otimes(ZZ')^{-1}ZP = \Sigma\otimes 0 = 0
\end{aligned}
$$

上式说明了 $\hat{\Theta}$ 与 $\hat{\varepsilon}$ 互不相关，定理得证。 ▮

注意到，若 $\operatorname{rk}Z < q+1$，由第一章定理 3.5 知 $Z'(ZZ')^{-}Z$ 是唯一的在 $\mathscr{L}(Z')$ 上的投影阵且秩为 $\operatorname{rk}Z$，如上证明可得

$$EQ = (n-\operatorname{rk}Z)\Sigma$$

下面我们将考虑预测问题有关的估计。

由最小二乘法我们已给出了 Θ 和 Σ 的估计，并建立了回归方程(2.18)式。现给自变量以新的值 $\boldsymbol{z}_0$，根据回归模型，有

$$\boldsymbol{x}_0 = \Theta\boldsymbol{z}_0 + \boldsymbol{\varepsilon}_0$$

其中 $\boldsymbol{x}_0$ 是相应于 $\boldsymbol{z}_0$ 的因变量观察向量，$\boldsymbol{\varepsilon}_0$ 是误差向量，且 $E\boldsymbol{\varepsilon}_0=\boldsymbol{o}$，$\operatorname{cov}(\boldsymbol{\varepsilon}_0)=\Sigma$，$\operatorname{cov}(\boldsymbol{\varepsilon}_0,\boldsymbol{\varepsilon}_j)=0$ $(j=1,2,\cdots,n)$，即 $\boldsymbol{\varepsilon}_0$ 与 $\boldsymbol{\varepsilon}$ 不相关。由预测方程可计算出在 $\boldsymbol{z}_0$ 之下其预测值为

$$\hat{\boldsymbol{x}}_0 = \hat{\Theta}\boldsymbol{z}_0$$

而预测误差为

$$\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0 = \boldsymbol{x}_0 - \hat{\Theta}\boldsymbol{z}_0$$

于是

$$E(\boldsymbol{x}_0-\hat{\boldsymbol{x}}_0) = E\boldsymbol{x}_0 - E\hat{\Theta}\boldsymbol{z}_0 = \Theta\boldsymbol{z}_0 - \Theta\boldsymbol{z}_0 = \boldsymbol{o}$$

这说明了 $\hat{\boldsymbol{x}}_0$ 是 $\boldsymbol{x}_0$ 的一个无偏预测量，而预测误差的协方差阵 $\tilde{\Sigma}=(\tilde{\sigma}_{ij})$ 可如下求出：

由于 $\boldsymbol{x}_0$ 与 X ($\boldsymbol{\varepsilon}_0$ 与 $\boldsymbol{\varepsilon}$) 互不相关，所以

$$\operatorname{cov}(\boldsymbol{x}_0-\hat{\boldsymbol{x}}_0) = \operatorname{cov}(\boldsymbol{x}_0-\hat{\Theta}\boldsymbol{z}_0) = \operatorname{cov}(\boldsymbol{x}_0) + \operatorname{cov}(\hat{\Theta}\boldsymbol{z}_0)$$

而

$$
\begin{aligned}
\operatorname{cov}(\hat{\Theta}\boldsymbol{z}_0) &= \operatorname{cov}((I\otimes\boldsymbol{z}_0')\operatorname{vec}\hat{\Theta}) = (I\otimes\boldsymbol{z}_0')\operatorname{cov}(\operatorname{vec}\hat{\Theta})(I\otimes\boldsymbol{z}_0)\\
&= (I\otimes\boldsymbol{z}_0')(\Sigma\otimes(ZZ')^{-1})(I\otimes\boldsymbol{z}_0) = \Sigma\otimes\boldsymbol{z}_0'(ZZ')^{-1}\boldsymbol{z}_0\\
&= \boldsymbol{z}_0'(ZZ')^{-1}\boldsymbol{z}_0\cdot\Sigma
\end{aligned}
$$

由于
$$\text{cov}(x_0) = \Sigma$$
所以有
$$\hat{\Sigma} = \text{cov}(x_0 - \hat{x}_0) = \Sigma \cdot (1 + z_0'(ZZ')^{-1}z_0) \quad (2.30)$$

（二） 极大似然估计及其性质

在(一)中讨论的 LS 估计，对模型没有作分布的假设。如果我们对模型作出正态分布的假设，此时可利用极大似然方法求得Θ和Σ的极大似然估计，并将看到Θ的 LS 估计与极大似然估计是一致的。

定理 2.3 在多对多的模型(1.12)中，设 $X \sim N_{p,n}(\Theta Z, \Sigma, I)$(或 $\varepsilon \sim N_{p,n}(0, \Sigma, I_n)$，$\text{rk}Z = q+1$，$n-(q+1) \geqslant p$，则$\Theta$和$\Sigma$的极大似然估计是
$$\hat{\Theta} = XZ'(ZZ')^{-1} \quad (2.31)$$
$$\hat{\Sigma} = \frac{1}{n}Q = \frac{1}{n}(XX' - \hat{\Theta}ZZ'\hat{\Theta}') \quad (2.32)$$

证明　当 $X \sim N_{p,n}(\Theta Z, \Sigma, I_n)$ 时，它的似然函数为
$$\begin{aligned}L(\Theta, \Sigma) &= (2\pi)^{-np/2}|\Sigma|^{-n/2} \\ &\quad \times \exp\left\{\text{tr}\left[-\frac{1}{2}\Sigma^{-1}(X-\Theta Z)(X-\Theta Z)'\right]\right\} \\ &= (2\pi)^{-np/2}|\Sigma|^{-n/2} \\ &\quad \times \exp\left\{\text{tr}\left[-\frac{1}{2}\Sigma^{-1}((X-\hat{\Theta}Z)(X-\hat{\Theta}Z)' \right.\right. \\ &\quad \left.\left. + (\hat{\Theta}-\Theta)ZZ'(\hat{\Theta}-\Theta)')\right]\right\}\end{aligned}$$
其中 $\hat{\Theta} = XZ'(ZZ')^{-1}$。

由于对非负定矩阵 $A = CC'$，有$\text{tr}A \geqslant 0$，所以
$$\text{tr}\Sigma^{-1}(\hat{\Theta}-\Theta)ZZ'(\hat{\Theta}-\Theta)' = \text{tr}CC' \geqslant 0 \quad (C = \Sigma^{-\frac{1}{2}}(\hat{\Theta}-\Theta)Z)$$
显然上式当 $\Theta = \hat{\Theta}$ 达到最小值 0，又由第四章引理1.1知，当$\Sigma = \hat{\Sigma} = \frac{1}{n}Q$ 时，$L(\hat{\Theta}, \Sigma)$ 达到最大，从而

$$L(\Theta,\Sigma) \leqslant L(\hat{\Theta},\hat{\Sigma})$$

即 $\hat{\Theta}$ 和 $\hat{\Sigma}$ 是 Θ 和 Σ 的极大似然估计. ▌

关于 $\hat{\Theta}$ 和 $\hat{\Sigma}=\frac{1}{n}Q$ 的分布有下面的定理.

定理 2.4 当 $X \sim N_{p,n}(\Theta Z,\Sigma,I_n)$, $\Sigma>0$, $n-(q+1)\geqslant p$, $\mathrm{rk}Z=q+1$, 则有下列结论:

(1) $\hat{\Theta}=XZ'(ZZ')^{-1} \sim N_{p,(q+1)}(\Theta,\Sigma,(ZZ')^{-1})$ (2.33)

(2) $Q=n\hat{\Sigma} \sim W_p(n-q-1,\Sigma)$ (2.34)

(3) $\hat{\Theta}$ 与 $\hat{\Sigma}$ 相互独立.

证明 记 $B=Z'(ZZ')^{-1}$, 按第二章定理 4.1,有

$$\hat{\Theta}=XB \sim N_{p,(q+1)}(\Theta ZB,\Sigma,B'B)$$

由于 $ZB=I_{q+1}$, $B'B=(ZZ')^{-1}$, 从而(2.33)式得证.

由 $Q=X(I_n-Z'(ZZ')^{-1}Z)X'=XPX'$, 其中 $P=I_n-Z'(ZZ')^{-1}Z$ 是投影阵,且 $\mathrm{rk}P=n-q-1$, 而 $\Theta ZP=0$, 按第三章推论 2.5 知道

$$Q \sim W_p(n-q-1,\Sigma)$$

从而(2.34)得证.

记 $G=XP$, 又 $\hat{\Theta}=XZ'(ZZ')^{-1}$, 由 P 的定义,有

$$P'\cdot I_n\cdot Z'(ZZ')^{-1}=0$$

根据第二章定理 4.2 便知道 G 与 $\hat{\Theta}$ 相互独立, 于是 $Q=GG'$ 与 $\hat{\Theta}$ 也相互独立. ▌

关于预测问题有关的估计在正态假设下有如下结论:

(1) $\boldsymbol{x}_0$ 的预测值 $\hat{\boldsymbol{x}}_0=\hat{\Theta}\boldsymbol{z}_0 \sim N_p(\Theta\boldsymbol{z}_0,\boldsymbol{z}_0'(ZZ')^{-1}\boldsymbol{z}_0\cdot\Sigma)$.(利用第二章定理 4.1 便可得出).

(2) 预测误差 $(\boldsymbol{x}_0-\hat{\boldsymbol{x}}_0) \sim N_p(0,(1+\boldsymbol{z}_0'(ZZ')^{-1}\boldsymbol{z}_0)\Sigma)$.(这是由于 $\boldsymbol{x}_0$ 和 $\hat{\boldsymbol{x}}_0=\hat{\Theta}\boldsymbol{z}_0$ 均为正态分布且相互独立).

例 2.1 要购买计算机的公司为了确定所需要的设备,公司的技术人员从同类的公司收集数据, 以建立计算机硬件设备的预测方程. 其数据如表 2.1,其中 z_1 = 顾客指令(以千条计), z_2 = 插

入与删除指令数(以千条计)，z_1 = 主机运行时间(小时)，z_2=磁盘输入/输出能力。

表 2.1　计算机数据

z_1	123.5	146.5	133.9	128.5	151.5	136.2	92.0
z_2	2.108	9.213	1.905	0.815	1.061	8.603	1.125
x_1	141.5	168.9	154.8	146.5	172.8	160.1	108.5
x_2	301.8	396.1	328.2	307.4	362.4	369.5	229.1

由表 2.1 不难写出设计矩阵和因变量观测矩阵为

$$Z=\begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1\\ 123.5 & 146.1 & 133.9 & 128.5 & 151.5 & 136.2 & 92.0\\ 2.108 & 9.213 & 1.905 & 0.815 & 1.061 & 8.603 & 1.125\end{pmatrix}$$

$$X=\begin{pmatrix}141.5 & 168.9 & 154.8 & 146.5 & 172.8 & 160.1 & 108.5\\ 301.8 & 396.1 & 328.2 & 307.4 & 362.4 & 369.5 & 229.1\end{pmatrix}$$

并计算得

$$(ZZ')^{-1}=\begin{pmatrix}8.17969 & -0.06411 & 0.08831\\ -0.06411 & 0.00052 & -0.00107\\ 0.08831 & -0.00107 & 0.01440\end{pmatrix}$$

由(2.7)式求得回归系数矩阵Θ的 LS 估计为

$$\hat{\Theta}=XZ'(ZZ')^{-1}=\begin{pmatrix}8.42 & 1.08 & 0.42\\ 14.42 & 2.25 & 5.67\end{pmatrix}$$

从而回归方程为

$$\hat{x}=\begin{pmatrix}\hat{x}_1\\ \hat{x}_2\end{pmatrix}=\hat{\Theta}z=\begin{pmatrix}8.42 & 1.08 & 0.42\\ 14.42 & 2.25 & 5.67\end{pmatrix}\begin{pmatrix}1\\ z_1\\ z_2\end{pmatrix}$$

$$=\begin{pmatrix}8.42+1.08z_1+0.42z_2\\ 14.42+2.25z_1+5.67z_2\end{pmatrix}$$

而Σ的极大似然估计为

$$\hat{\Sigma} = \frac{1}{n}(XX' - \hat{\Theta}ZZ'\hat{\Theta}) = \frac{1}{7}\begin{pmatrix} 5.80 & 5.30 \\ 5.30 & 13.13 \end{pmatrix}$$

$$= \begin{pmatrix} 0.8286 & 0.7571 \\ 0.7571 & 1.8757 \end{pmatrix}.$$

注意到在上述讨论的多对多回归模型中 $\Theta = (\theta_0\theta_1\cdots\theta_q)$，今后我们在对回归系数的检验和估计中更关心的是自变量 z 所对应的系数 $(\theta_1\ \theta_2\cdots\theta_q)$，为了更方便地求出 $(\theta_1\ \theta_2\cdots\theta_q)$ 的估计，我们把多对多模型改写成为

$$\begin{cases} X = \Theta\begin{pmatrix} \mathbf{1}'_n \\ \tilde{Z} \end{pmatrix} + \varepsilon \\ \varepsilon \sim N_{p,n}(0, \Sigma, I_n) \end{cases}$$

其中 $\Theta = (\theta_0\ \Theta^*)$，$\Theta^* = (\theta_1\ \theta_2\cdots\theta_q)$，$\tilde{Z} = (Z_1\cdots Z_n)$（注意到这里 $\tilde{Z}$ 与(1.12)中的 Z 的差别），$\mathbf{1}'_n = (1, 1, \cdots, 1)$。令

$$Z^* = \tilde{Z} - \bar{z}\mathbf{1}'_n = (Z_1 - \bar{z}, \cdots, Z_n - \bar{z}) = \tilde{Z}P \quad (2.35)$$

$$\alpha_0 = \theta_0 + \Theta^*\bar{z} \quad (2.36)$$

其中 $\bar{z} = \frac{1}{n}\sum_{j=1}^{n} Z_j$，$P = I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n$，这时上述模型可写成为

$$\begin{cases} X = (\alpha\Theta^*)\begin{pmatrix} \mathbf{1}'_n \\ Z^* \end{pmatrix} + \varepsilon \\ \varepsilon \sim N_{p,n}(0, \Sigma, I_n) \end{cases} \quad (2.37)$$

假设 $\text{rk}\ Z^* = q$，注意到 $Z^*\mathbf{1}_n = \tilde{Z}P\mathbf{1}_n = 0$，这时 $(\alpha\Theta^*)$ 的 LS 估计为

$$(\hat{\alpha}\hat{\Theta}^*) = X(\mathbf{1}_n Z^{*\prime})\begin{pmatrix} \mathbf{1}'_n\mathbf{1}_n & 0 \\ 0 & Z^*Z^{*\prime} \end{pmatrix}^{-1}$$

$$= \left(\frac{1}{n}X\mathbf{1}_n \quad XZ^{*\prime}(Z^*Z^*)^{-1}\right)$$

即

$$\hat{a} = \frac{1}{n} X\mathbf{1}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \tag{2.38}$$

$$\hat{\Theta}^* = XZ^{*\prime}(Z^*Z^{*\prime})^{-1} \tag{2.39}$$

由(2.36)可得

$$\hat{\theta}_0 = \bar{x} - \hat{\Theta}^*\bar{z} \tag{2.40}$$

显然(2.39)式的 $\hat{\Theta}^*$ 是下列方程的解

$$L_{ZZ}\Theta^* = L_{XZ} \tag{2.41}$$

其中

$$L_{ZZ} = Z^*Z^{*\prime},\quad L_{XZ} = XZ^{*\prime} \tag{2.42}$$

由第二章定理 4.1 知

$$\hat{\Theta}^* \sim N_{p\times q}(\Theta^*, \Sigma, (Z^*Z^{*\prime})^{-1}) \tag{2.43}$$

此时残差交叉乘积矩阵(2.22)可表为

$$\begin{aligned} Q &= \left[X - (\hat{a}\hat{\Theta}^*)\begin{pmatrix} \mathbf{1}_n' \\ Z^* \end{pmatrix}\right]\left[X - (\hat{a}\hat{\Theta}^*)\begin{pmatrix} \mathbf{1}_n' \\ Z^* \end{pmatrix}\right]' \\ &= XPX' - XZ^{*\prime}(Z^*Z^{*\prime})^{-1}Z^*X' \\ &= X[P - P\tilde{Z}'(\tilde{Z}P\tilde{Z}')^{-1}\tilde{Z}P]X' = XRX' \end{aligned} \tag{2.44}$$

其中

$$R = P - P\tilde{Z}'(\tilde{Z}P\tilde{Z}')^{-1}\tilde{Z}P \tag{2.45}$$

容易证明 R 是幂等对称阵，且 $\mathrm{rk}\, R = n - q - 1$，所以 $Q \sim W_p(n - q - 1, \Sigma)$，且与 $(\hat{a}\hat{\Theta}^*)$ 相互独立，于是由(2.43)式，有

$$(\hat{\Theta}^* - \Theta)(Z^*Z^{*\prime})^{-1}(\hat{\Theta}^* - \Theta^*)' \sim W_p(q, \Sigma)$$

且与 Q 独立。此时如果 $p = 1$，$q = 1$，则有

$$\|\hat{\theta}_1 - \theta_1\|^2 \cdot \sum_{i=1}^{n}(z_i - \bar{z})^2 \Big/ \frac{1}{n-2} Q \sim F(1, n-2) \tag{2.46}$$

*(三) 关于 Θ 的线性函数的估计

前面对回归系数矩阵 Θ 和协方差矩阵 Σ 的估计进行了讨论。然而在 ZZ' 不是非奇异矩阵时往往会带来一些问题，而在一些具体问题中有时只要求对 Θ 的某个线性函数进行估计，下面将对这个问题作一些探讨。

仍然考虑多对多回归模型(1.12).

定义 2.1 设Θ元素的线性函数

$$\rho = \operatorname{tr}\Theta A = \sum_{i=1}^{p}\sum_{j=1}^{q+1} a_{ji}\theta_{ij}$$

如果存在X元素的线性函数 $\operatorname{tr}XB$，使得

$$E\operatorname{tr}XB = \operatorname{tr}\Theta A = \rho$$

则称ρ是可估的，$\operatorname{tr}XB$ 称为ρ的线性无偏估计. 如果在ρ的线性无偏估计中存在方差最小的估计，则称它是ρ的最优线性无偏估计或者马尔科夫估计.

引理 2.1 记号如上，则

(1) $E\operatorname{tr}XB = \operatorname{tr}\Theta ZB$

(2) $\operatorname{var}(\operatorname{tr}XB) = \operatorname{tr}B\Sigma B'$

(3) $\operatorname{cov}(\operatorname{tr}XB, \operatorname{tr}XD) = \operatorname{tr}B\Sigma D'$

证明 (1) 是显然的.

(2) 记 $B = (\boldsymbol{b}_1 \cdots \boldsymbol{b}_n)'$，于是

$$\operatorname{tr}XB = \operatorname{tr}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{b}_i' = \sum_{i=1}^{n}\boldsymbol{b}_i'\boldsymbol{x}_i$$

由于$\boldsymbol{x}_i$互不相关且有相同的协方差矩阵Σ，从而

$$\operatorname{var}(\operatorname{tr}XB) = \operatorname{var}\left(\sum_{i=1}^{n}\boldsymbol{b}_i'\boldsymbol{x}_i\right) = \sum_{i=1}^{n}\boldsymbol{b}_i'\Sigma\boldsymbol{b}_i$$

$$= \sum_{i=1}^{n}\operatorname{tr}\Sigma\boldsymbol{b}_i\boldsymbol{b}_i' = \operatorname{tr}\Sigma B'B = \operatorname{tr}B\Sigma B'$$

(3)的证明与(2)类似. ∎

定理 2.5 $\rho = \operatorname{tr}\Theta A$ 可估的充分必要条件是存在矩阵 B 使得 $ZB = A$ 成立.

证明 $\rho = \operatorname{tr}\Theta A$ 可估的充分必要条件是存在B使得

$$\rho = \operatorname{tr}\Theta A = E\operatorname{tr}XB = \operatorname{tr}\Theta ZB$$

对一切Θ成立，于是从 $\operatorname{tr}\Theta(A - ZB) = 0$ 得到

$$ZB = A.$$ ∎

下面记 $\hat{\Theta}=XZ'(ZZ')^-$ 是Θ的任一 LS 估计（参看 (2.9) 式）.

定理 2.6 对可估参数 $\rho=\mathrm{tr}\Theta A$ 而言，有

(1) $\mathrm{tr}\hat{\Theta}A$ 是唯一确定的且 $E\,\mathrm{tr}\hat{\Theta}A=\mathrm{tr}\Theta A$

(2) $\mathrm{var}(\mathrm{tr}\hat{\Theta}A)=\mathrm{tr}(ZZ')^+A\Sigma A'$.

证明 由于 $\rho=\mathrm{tr}\Theta A$ 可估，存在 B 使得 $A=ZB$. 而 $\hat{\Theta}A=XZ'(ZZ')^-ZB$, $Z'(ZZ')^-Z$ 是 $\mathscr{L}(Z')$ 上投影阵，与广义逆无关，因而 $\hat{\Theta}A=XZ'(ZZ')^+A$ 唯一确定，(1)便得证. 而

$$E\,\mathrm{tr}\hat{\Theta}A=\mathrm{tr}\Theta ZZ'(ZZ')^-ZB=\mathrm{tr}\Theta ZB=\mathrm{tr}\Theta A$$

由引理 2.1 可计算得

$$\begin{aligned}\mathrm{var}(\mathrm{tr}\hat{\Theta}A)&=\mathrm{var}(\mathrm{tr}XZ'(ZZ')^+A)\\&=\mathrm{tr}[Z'(ZZ')^+A\cdot\Sigma\cdot A'(ZZ')^+Z]\\&=\mathrm{tr}[(ZZ')^+A\Sigma A']\end{aligned}$$

(2)便得证. ▌

定理 2.7(Gauss-Markov 定理) 设 $\rho=\mathrm{tr}\Theta A$ 可估，则 $\mathrm{tr}\hat{\Theta}A$ 是 ρ 的唯一的最优线性无偏估计.

证明 由定理 2.6 知 $\mathrm{tr}\hat{\Theta}A$ 是 $\mathrm{tr}\Theta A$ 的无偏估计且它的方差为 $\mathrm{tr}(ZZ')^+A\Sigma A'$. 由于 $\rho=\mathrm{tr}\Theta A$ 可估，故对它的任一无偏估计 $\mathrm{tr}\,XB$ 有 $ZB=A$, 于是

$$\begin{aligned}\mathrm{var}(\mathrm{tr}\,XB)&=\mathrm{tr}\,B\Sigma B'=\mathrm{tr}\Sigma^{\frac{1}{2}}B'B\Sigma^{\frac{1}{2}}\\&\geqslant\mathrm{tr}\Sigma^{\frac{1}{2}}B'Z'(ZZ')^+ZB\Sigma^{\frac{1}{2}}\\&=\mathrm{tr}(ZZ')^+A\Sigma A'=\mathrm{var}(\mathrm{tr}\hat{\Theta}A)\end{aligned}$$

按定义 2.1, $\mathrm{tr}\hat{\Theta}A$ 是 ρ 的最优线性无偏估计.

下面证明唯一性. 记 $\hat{\rho}_1=\mathrm{tr}\hat{\Theta}A$, 另一个最优线性无偏估计记为 $\hat{\rho}_2=\mathrm{tr}\,XB$, 则 $A=ZB$. 于是由引理 2.1，有

$$\begin{aligned}\mathrm{cov}(\hat{\rho}_1,\hat{\rho}_2)&=\mathrm{cov}(\mathrm{tr}\,XB,\mathrm{tr}XZ'(ZZ')^+ZB)\\&=\mathrm{tr}B\Sigma B'Z'(ZZ')^+Z=\mathrm{tr}(ZZ')^+A\Sigma A'\end{aligned}$$

即 $\mathrm{var}(\hat{\rho}_1)=\mathrm{var}(\hat{\rho}_2)=\mathrm{cov}(\hat{\rho}_1,\hat{\rho}_2)$, 从而

$$E\hat{\rho}_1=E\hat{\rho}_2\quad E(\hat{\rho}_1-\hat{\rho}_2)^2=\mathrm{var}(\hat{\rho}_1)+\mathrm{var}(\hat{\rho}_2)-2\mathrm{cov}(\hat{\rho}_1,\hat{\rho}_1)=0$$

即 $\hat{\rho}_1=\hat{\rho}_2$ 以概率 1 成立. 定理得证. ▌

定理2.8 设 $X\sim N_{p,n}(\Theta Z,\Sigma,I_n)$, $\rho=\mathrm{tr}\Theta A$ 可估, 则 ρ 的最优线性无偏估计

$$\hat{\rho}=\mathrm{tr}\hat{\Theta}A\sim N(\mathrm{tr}\Theta A,\mathrm{tr}(ZZ')^{+}A\Sigma A') \tag{2.47}$$

证明 $\hat{\rho}=\mathrm{tr}\hat{\Theta}A=\mathrm{tr}\,XZ'(ZZ')^{+}A$ 是联合正态变量 x_{ij} 的线性组合,仍服从正态分布. 其期望、方差由定理 2.7 确定故(2.47)式成立. ▌

对于 Θ 而言,仅当每一元素 θ_{ij} 可估时才说 Θ 是可估的. 而对于可估函数 $\rho=\mathrm{tr}\Theta A$ 而言,尽管当 ZZ' 不满秩时 LS 估计不确定,但它仍有最优线性无偏估计 $\mathrm{tr}\hat{\Theta}A$.

最后我们指出,多元线性模型也可通过把 X 改写为 vec X 后成为一元的广义线性模型来处理. (1.12) 可写为:

$$\begin{cases}\mathrm{vec}\,X=(I\otimes Z')\mathrm{vec}\,\Theta+\mathrm{vec}\,\epsilon\\ E(\mathrm{vec}\,\epsilon)=0,\ \mathrm{cov}(\mathrm{vec}\,\epsilon)=\Sigma\otimes I_n\end{cases}$$

从而一元线性模型的结论都可在多元线性模型中适用, 读者可自行比较得到.

§6.3 多元正态线性模型中回归参数的假设检验

在一元统计的回归分析中, 我们考虑了 q 个自变量与一个因变量之间是否存在显著线性关系的问题, 也考虑了每个自变量对因变量是否有显著影响的问题. 对于多对多的线性回归模型也要讨论这些问题,即要对回归系数进行各种检验.

设 $X=(x_1\cdots x_n)\sim N_{p,n}(\Theta Z,\Sigma,I_n)$, 其中 Θ,Z 分别由 (1.9) 及(1.10)式所定义, 并设 ZZ' 非奇异且 $n\geqslant p+q+1$.

(一) 关于 Θ 的子矩阵的检验

设 $\Theta=(\Theta_1,\Theta_2)$, 其中 Θ_1 是 $p\times(q+1-s)$ 阶, Θ_2 是 $p\times s$ 阶, $(s<q+1)$. 考虑检验如下假设

$$H_0:\Theta_2=0,\ H_1:\Theta_2\neq 0 \tag{3.1}$$

即检验因变量是否依赖于自变量 $z_{q+1-s},\cdots,z_q$.

我们利用似然比方法导出检验统计量．此时参数空间为

$$\Omega=\{(\Theta,\Sigma):\ \Theta \text{ 为 } p\times(q+1) \text{ 阶矩阵},\ \Sigma>0\}$$

似然函数为

$$L(\Theta,\Sigma)=(2\pi)^{-np/2}|\Sigma|^{-n/2}$$
$$\times\exp\left\{\operatorname{tr}\left[-\frac{1}{2}\Sigma^{-1}(X-\Theta Z)(X-\Theta Z)'\right]\right\}$$

按定理 2.3 的证明，知当

$$\Theta=\hat{\Theta}=XZ'(ZZ')^{-1}$$

和

$$\Sigma=\hat{\Sigma}=\frac{1}{n}Q_{\Omega}=\frac{1}{n}(X-\hat{\Theta}Z)(X-\hat{\Theta}Z)'$$
$$=\frac{1}{n}X[I_n-Z'(ZZ')^{-1}Z]X'$$

时，$L(\Theta,\Sigma)$ 达到最大，即

$$\max_{\Omega} L(\Theta,\Sigma)=L(\hat{\Theta},\hat{\Sigma})$$
$$=(2\pi)^{-np/2}\left|\frac{1}{n}(X-\hat{\Theta}Z)(X-\hat{\Theta}Z)'\right|^{-n/2}$$
$$\times\exp\left(-\frac{1}{2}np\right)$$

在 H_0 之下参数空间为 $\Omega_0=\{(\Theta,\Sigma):\ \Sigma>0,\ \Theta_2=0\}$ 时，记 $Z=\begin{pmatrix}Z_1\\Z_2\end{pmatrix}$，其中 Z_1 是$(q+1-s)\times n$ 阶，Z_2 是 $s\times n$ 阶。这时模型为

$$X\sim N_{p,n}(\Theta_1Z_1,\Sigma,I_n)$$

似然函数为

$$L(\Theta_1,\Sigma)=(2\pi)^{-np/2}|\Sigma|^{-n/2}$$
$$\times\exp\left\{\operatorname{tr}\left[-\frac{1}{2}\Sigma^{-1}(X-\Theta_1Z_1)(X-\Theta_1Z_1)'\right]\right\}$$

同上可知当

$$\Theta_1 = \hat{\Theta}_1 = XZ_1'(Z_1Z_1')^{-1}$$

$$\Sigma = \hat{\Sigma}_0 = \frac{1}{n} Q_{\Omega_0} = \frac{1}{n}(X - \hat{\Theta}_1 Z_1)(X - \hat{\Theta}_1 Z_1)'$$

$$= \frac{1}{n} X[I_n - Z_1'(Z_1Z_1')^{-1}Z_1]X'$$

时达到最大，即

$$\max_{\Omega_0} L(\Theta, \Sigma) = L(\hat{\Theta}_1, \hat{\Sigma}_0)$$

$$= (2\pi)^{-np/2} \left| \frac{1}{n}(X - \hat{\Theta}_1 Z_1)(X - \hat{\Theta}_1 Z_1)' \right|^{-n/2}$$

$$\times \exp\left(-\frac{np}{2}\right)$$

于是似然比统计量为

$$\Lambda^* = \max_{\Omega_0} L(\Theta, \Sigma) / \max_{\Omega} L(\Theta, \Sigma) = (|\hat{\Sigma}_0| / |\hat{\Sigma}|)^{-n/2} \quad (3.2)$$

它等价于下列检验统计量

$$\Lambda = |Q_\Omega| / |Q_{\Omega_0}| = |\hat{\Sigma}| / |\hat{\Sigma}_0| \quad (3.3)$$

其中

$$Q_\Omega = X[I - Z'(ZZ')^{-1}Z]X' = n\hat{\Sigma}$$

$$Q_{\Omega_0} = X[I - Z_1'(Z_1Z_1')^{-1}Z_1]X' = n\hat{\Sigma}_0$$

对于给定的显著水平 α，当 Λ 小于相应的临界值时，就放弃（3.1）中的原假设 H_0。即得到如下定理：

定理 3.1 设 $X = (x_1 \cdots x_n) \sim N_{p,n}(\Theta Z, \Sigma, I_n)$ 且 x_i 相互独立，$\Theta = (\Theta_1, \Theta_2)$，其中 Θ_1 是 $p \times (q + 1 - s)$ 阶，Θ_2 是 $p \times s$ 阶，则拒绝(3.1)中的原假设 H_0 的拒绝域是由

$$\Lambda = |Q_\Omega| / |Q_{\Omega_0}| \leqslant \Lambda_\alpha \quad (3.4)$$

给出，其中 Λ_α 满足 $P(\Lambda \leqslant \Lambda_\alpha) = \alpha$（$\alpha$ 是给定的显著水平）。

定理 3.2 条件同定理 3.1，则当(3.1)中 H_0 为真时，检验统计量(3.3)为 Wilks 统计量，即

$$\Lambda \sim \Lambda(p, n - q - 1, s) \quad (3.5)$$

证明　由(3.2)和(3.3)可知

$$Q_\Omega = X(I-P_1)X', \quad Q_{\Omega_0} = X(I-P_2)X'$$

其中 P_1 是 $\mathscr{L}(Z')$ 上投影阵，P_2 是 $\mathscr{L}(Z_1')$ 上的投影阵，且 $Z'=(Z_1'Z_2')$，由第一章定理 3.7 知道 $P_1=P_2+R$，且 $P_2R=0$，其中 R 是 $\mathscr{L}[(I-P_2)Z_2']$ 上的投影矩阵。所以

$$Q_{\Omega_0} = X(I-P_2)X' = X(I-P_1)X' + XRX' = Q_\Omega + XRX'$$

且

$$(I-P_1)R = (I-P_2-R)R = 0 \quad \mathrm{rk}(I-P_1) = n-q-1,$$

$$\mathrm{rk}(I-P_2) = n-(q+1-s), \quad \mathrm{rk}\, R = s$$

以及在 H_0 成立时

$$\Theta Z(I-P_1) = 0 \quad \Theta_1 Z_1(I-P_2) = 0$$

由第三章定理 2.10 知道在 H_0 成立时

$$Q_\Omega \sim W_p(n-q-1,\Sigma) \quad XRX' \sim W_p(s,\Sigma)$$

且 Q_Ω 与 XRX' 相互独立。所以按 Wilks Λ 统计量的定义知道

$$\Lambda = |Q_\Omega|/|Q_{\Omega_0}| = |Q_\Omega|/|Q_\Omega + XRX'|$$

$$\sim \Lambda(p,n-q-1,s) \quad ▌$$

注：在(3.1)中，取 $\Theta_2=(\theta_1,\cdots\theta_q)$，即它是全体自变量 $z_1,\cdots,z_q$ 所对应的系数矩阵，如果检验结果是 $H_0:\Theta_2=0$ 成立，那么我们认为自变量 $z_1,\cdots,z_q$ 与因变量 $x_1,\cdots,x_p$ 之间的关系不符合线性模型，此时应考虑其它模型。

类似地，对于指定的 $p\times s$ 阶矩阵 Θ_2^*，要检验假设

$$H_0:\Theta_2=\Theta_2^* \qquad H_1:\Theta_2\neq\Theta_2^* \tag{3.6}$$

此时我们只须令 $Y=X-\Theta_2^*Z_2$，由于

$$X \sim N_{p,n}(\Theta_1Z_1+\Theta_2Z_2,\Sigma,I_n)$$

所以

$$Y \sim N_{p,n}(\Theta_1Z_1+\tilde{\Theta}_2Z_2,\Sigma,I_n)$$

其中 $\tilde{\Theta}_2=\Theta_2-\Theta_2^*$，此时检验(3.6)就等价于检验假设

$$H_0':\tilde{\Theta}_2=0, \qquad H_1':\tilde{\Theta}_2\neq 0$$

这正是检验假设(3.1)的问题。因此检验假设(3.6)的检验统计量为

$$\Lambda = |\tilde{Q}_{\Omega}|/|\tilde{Q}_{\Omega_0}|$$

其中

$$\tilde{Q}_{\Omega} = (X - \Theta_2^* Z_2)[I - Z'(ZZ')^{-1}Z]^{-1}(X - \Theta_2^* Z_2)'$$

$$\tilde{Q}_{\Omega_0} = (X - \Theta_2^* Z_2)[I - Z_1'(Z_1 Z_1')^{-1}Z_1]^{-1}(X - \Theta_2^* Z_2)'$$

注：上述证明中我们假设 ZZ' 非奇异、事实上，由于 Q_{Ω} 和 Q_{Ω_0} 与 $(ZZ')^-$ 无关而且是唯一确定的，从而当 $\mathrm{rk}Z < q+1$ 时也是正确的，只要作一些小改动，把 $q+1$ 改为 $q'+1 = \mathrm{rk}Z$，s 改为 $s' = \mathrm{rk}(Z_2)$，于是 $\Lambda \sim \Lambda(p, n-q'-1, s')$.

更一般地，对于 $m \times p$ 阶矩阵 B，$\mathrm{rk}\, B = m$；$(q+1)\times t$ 阶矩阵 D；$m \times t$ 阶矩阵 K。我们要检验假设

$$H_0: B\Theta D = K, \qquad H_1: B\Theta D \neq K \tag{3.7}$$

当然这里要求 $B\Theta D = K$ 应该是相容的.显然，当 $B = I_p$，$D = \begin{pmatrix} 0 \\ I_s \end{pmatrix}$ 时，(3.7)便是前面所讲的假设形式．检验(3.7)的统计量可以用似然比方法求出，但较复杂，我们用另一种方法说明似然比统计量，而且直观意义明确．由定理 2.4 知

$$Q = X[I - Z'(ZZ')^{-1}Z]X' \sim W_p(n-q-1, \Sigma)$$

$$\hat{\Theta} = XZ'(ZZ')^{-1} \sim N_{p,q+1}(\Theta, \Sigma, (ZZ')^{-1})$$

且 Q 与 $\hat{\Theta}$ 相互独立．由第二章定理 4.1 以及第三章定理 2.7 知

$$U = B\hat{\Theta}D \sim N_{m,t}(B\Theta D, B\Sigma B', D'(ZZ')^{-1}D)$$

$$V = BQB' \sim W_m(n-q-1, B\Sigma B') \tag{3.8}$$

且 U 与 V 相互独立。于是在 H_0 成立时

$$B\hat{\Theta}D - K \sim N_{m,t}(0, B\Sigma B', D'(ZZ')^{-1}D)$$

利用第三章推论 2.4

$$W = (B\hat{\Theta}D - K)[D'(ZZ')^{-1}D]^{-1}(B\hat{\Theta}D - K)' \sim W_m(t, B\Sigma B') \tag{3.9}$$

仍与 V 独立。所以可取检验统计量

$$\Lambda = |V|/|V+W| \sim \Lambda(m, n-q-1, t) \tag{3.10}$$

当 $\Lambda \leqslant \Lambda_\alpha$ 时，拒绝原假设 $H_0: B\Theta D = K$，其中 $P(\Lambda \leqslant \Lambda_\alpha) =$

α. 从而检验(3.7)的步骤如下:

(1) 计算 $\hat{\Theta} = XZ'(ZZ')^{-1}$ 和 $Q = X(I - Z'(ZZ')^{-1}Z)X'$ (或先计算 Z^*, 再由(2.39)、(2.40)及(2.44)求出 $\hat{\theta}^*, \hat{\theta}_0$ 和 Q)

(2) 计算 $B\hat{\Theta}D$ 和 $V = BQB'$

(3) 计算 $W = (B\hat{\Theta}D - K)[D'(ZZ')^{-1}D]^{-1}(B\hat{\Theta}D - K)'$

(4) 计算 $\Lambda = |V|/|V + W|$

(5) 对于给定的显著水平 α, 求出 Λ_α 使得 $P(\Lambda(m, n-q-1, t) \leqslant \Lambda_\alpha) = \alpha$. 这里当 m, t 为 1,2 时可利用第三章表 3.1 查 F 表, 当 $m > 3$, $t > 3$ 时, 可利用第三章公式(3.6), (3.7)(视 n 的大小)查 χ^2 表。

(6) 根据所得临界值来判断是否拒绝 H_0。

最后我们要指出, 在定理 3.2 的证明中 XRX' 的直观意义当时并不明确。事实上可以验证, 若取

$$D = \begin{pmatrix} 0 \\ I_t \end{pmatrix}\begin{matrix} q+1-t \\ t \end{matrix}, \quad \hat{\Theta} = (\hat{\Theta}_1 \hat{\Theta}_2)$$

则有

$$\hat{\Theta}_2 = \hat{\Theta} \cdot D$$

$$XRX' = \hat{\Theta}_2[D(ZZ')^{-1}D']\hat{\Theta}_2'$$

(留作习题), 因此在实际运用中有时可以直接从 Q 和 $\hat{\Theta}$ 着手。

例 3.1 在例 2.1 中建立的线性回归方程中, 检验自变 z_1, z_2 对因变量 x_1 和 x_2 有无显著影响, 再检验 z_2 对 x_1 和 x_2 有无显著影响。

此时 $p = 2$, $q + 1 = 3$, $n = 7$. $\Theta = (\theta_0 \theta_1 \theta_2)$ 是 2×3 阶矩阵。

检验 z_1 和 z_2 对 x_1 和 x_2 是否有显著影响, 即检验

$$H_0: \Theta_2 = (\theta_1 \theta_2) = 0 \quad H_1: \Theta_2 \neq 0$$

按定理 3.2 的方法, 由例 2.1 已求出了 $\hat{\Sigma}$, 即

$$\hat{\Sigma} = \begin{pmatrix} 0.8286 & 0.7571 \\ 0.7571 & 1.8757 \end{pmatrix}$$

而 $\hat{\Sigma}_0 = (X - \hat{\Theta}_1 Z_1)(X - \hat{\Theta}_1 Z_1)'$, $Z_1 = (1, \cdots, 1)_{1\times 7}$. 又

$$\hat{\Theta}_1 = XZ_1'(Z_1Z_1')^{-1} = (150.44 \quad 327.79)$$

所以

$$\hat{\Sigma}_0 = \begin{pmatrix} 401.068 & 987.0477 \\ 987.0477 & 2633.5641 \end{pmatrix}$$

根据(3.3)式可计算得

$$\Lambda = |\hat{\Sigma}| / |\hat{\Sigma}_0| = \frac{0.9801}{81975.56} = (0.00346)^2$$

而

$$\Lambda \sim \Lambda(2, 7-2-1, 2) = \Lambda(2, 4, 2)$$

根据第三章表 3.1 知

$$F = \frac{3}{2} \cdot \frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F(4, 6)$$

由 Λ 的值可计算得

$$F = \frac{3}{2} \cdot \frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} = \frac{3}{2} \cdot \frac{1-0.00346}{0.00346} = 432.03.$$

对于给定 $\alpha = 0.05$, 查 F 表得临界值 $F_{0.95}(4, 6) = 4.53$. 显然 $432.03 > 4.53$,故拒绝 H_0(Λ 越小, 则 $(1-\sqrt{\Lambda})/\sqrt{\Lambda}$ 越大), 即 $z_1 z_2$ 对 x_1, x_2 有显著影响.

检验 z_2 对 x_1, x_2 是否有显著影响,即检验

$$H_0: \Theta_2 = \theta_2 = 0 \qquad H_1: \theta_2 \neq 0$$

此时利用统计量(3.10)更为方便. 取 $B = I_2$, $D' = (0\ 0\ 1)$, 根据(3.8)和(3.9)式可计算得

$$V = BQB' = I_2 Q I_2' = Q = \begin{pmatrix} 5.80 & 5.30 \\ 5.30 & 13.13 \end{pmatrix}$$

$$W = (B\hat{\Theta}D)[D'(ZZ')^{-1}D]^{-1}(B\hat{\Theta}D)'$$

$$= \begin{pmatrix} 0.42 \\ 5.67 \end{pmatrix} (0.0144)^{-1}(0.42 \ \ 5.67)$$

$$= \begin{pmatrix} 12.25 & 165.375 \\ 165.375 & 2232.5625 \end{pmatrix}$$

从而

$$\Lambda = |V|/|V+W| = \frac{48.064}{236.978} = (0.0649)^2$$

再计算

$$F = \frac{3}{1} \cdot \frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} = 43.225$$

查 F 表得 $F_{0.95}(2.6) = 5.14$，显然 $F > 5.14$，故拒绝 H_0，即 z_2 对 x_1, x_2 有显著影响。

例 3.2 表 3.1 中的数据是哥尔菲大学动物系一次试验中取得的。在 25 桶鲜鱼保存液中混有不同剂量的铜 z_{1j}（每立升毫克），并记录下每桶鱼的平均重量 z_{2j} ($j = 1, 2, \cdots, 25$)，也记录下在 8，14，24，36，48 小时后死亡的鱼的比例数。为了稳定变差，特性（比例）变量 p_i 将由反正弦函数变换成为 $x_i = \arcsin\sqrt{p_i}$ ($i = 1, \cdots 5$)，回归模型为

$$X_{5\times 25} = \Theta_{5\times 3} Z_{3\times 25} + \varepsilon$$

利用计算机求得回归系数 Θ 的 LS 估计为

$$\hat{\Theta} = XZ'(ZZ')^{-1} = \begin{pmatrix} -94.601 & 19.275 & -24.683 \\ -229.176 & 43.667 & -20.027 \\ -258.803 & 48.990 & 3.332 \\ -246.048 & 47.327 & 5.224 \\ -232.958 & 45.098 & 8.239 \end{pmatrix}$$

另外还求得 $(x_1, \cdots, x_5)$ 的相关矩阵为

$$R = \begin{pmatrix} 1 & 0.353 & -0.124 & -0.334 & -0.322 \\ & 1 & 0.354 & 0.309 & 0.322 \\ & & 1 & 0.916 & 0.889 \\ & & & 1 & 0.986 \\ & & & & 1 \end{pmatrix}$$

表 3.1 对 25 桶鲜鱼的试验数据

x'					z'		
x_1	x_2	x_3	x_4	x_5	z_0	z_1（剂量对数）	z_2（平均重量）
0	0	30	30	30	1	5.5984	0.6695
0	18.435	33.211	33.211	33.211	1	6.0161	0.6405
0	45	60	71.565	71.565	1	6.4134	0.729
22.786	53.729	90	90	90	1	6.8459	0.77
42.130	90	90	90	90	1	7.2793	0.5655
0	12.921	26.565	26.565	26.565	1	5.5984	0.782
12.921	18.435	33.211	33.211	33.211	1	6.0161	0.812
12.921	42.130	77.079	90	90	1	6.4135	0.8125
18.435	56.789	90	90	90	1	6.8459	0.869
26.565	67.214	90	90	90	1	7.2793	0.8395
0	0	0	0	12.921	1	5.5984	0.8615
0	12.921	22.786	30	33.211	1	6.0161	0.9045
0	22.786	77.079	77.079	77.079	1	6.4135	1.028
0	47.869	77.079	90	90	1	6.8459	1.0445
18.435	67.213	90	90	90	1	7.2793	1.0455
0	0	0	12.921	18.435	1	5.5984	0.6195
0	12.921	22.786	26.565	30	1	6.0161	0.5305
18.435	42.130	77.079	77.079	77.079	1	6.4135	0.597
18.435	56.789	90	90	90	1	6.8459	0.6385
36.271	77.079	90	90	90	1	7.2793	0.6645
0	12.921	26.565	26.565	26.565	1	5.5984	0.5685
0	0	22.786	30	30	1	6.0162	0.604
0	39.231	71.565	90	90	1	6.4135	0.6325
12.921	53.729	90	90	90	1	6.8459	0.6845
33.211	67.213	90	90	90	1	7.2793	0.728

从 $\hat{\Theta}$ 与 R 可以观察到 x_3，x_4，x_5 之间存在着高度相关且回归系数很接近的。这是由于鱼对有毒物质的适应性。但我们感兴趣的是桶中鱼的平均重量是否对死亡速度有影响。即要考察 z_2（平均重量）对因变量 $x_i(i=1,\cdots,5$，不同的 i 对应着不同的死亡时间）是否有显著影响。对于回归系数矩阵 $\Theta=(\theta_0\theta_1\theta_2)$，我们要检

验

$$H_0: B\Theta D = 0 \qquad H_1: B\Theta D \neq 0$$

其中 $B = I_5$,$D = (0\ 0\ 1)'$, 此时 $m = 5$, $s = 1$, $q + 1 = 3$, $n = 25$. 类似于上例计算得到

$$\Lambda = |V|/|V + W| = 0.66 \quad (\Lambda \sim \Lambda(5,22,1))$$

因而

$$F = \frac{18}{5} \cdot \frac{1 - \Lambda}{\Lambda} = 1.90 \quad (F \sim F(5,18))$$

对于显著水平 $\alpha = 0.05$ 和自由度(5,18)查 F 表得 $F_{0.95}(5,18) = 2.77$, 显然 $F = 1.90 < 2.77$. 因此可以接受 H_0, 即鱼的平均重量对死亡速度没有显著影响.

(二) 多元回归的预测问题

在多对多回归模型(1.12)中, 设 $X \sim N_{p,n}(\Theta Z, \Sigma, I_n)$, 并且建立了回归方程

$$\boldsymbol{x} = \hat{\Theta}\boldsymbol{z} \tag{3.11}$$

现任给自变量以值 $\boldsymbol{z}_0$, 要预测相应的 $\boldsymbol{x}_0$. 显然, 由回归方程 (3.11) 可得预测值向量 $\hat{\boldsymbol{x}}_0 = \hat{\Theta}\boldsymbol{z}_0$. 下面我们要考虑 $\Theta\boldsymbol{z}_0$ 的 $1 - \alpha$ 置信区域及 $\boldsymbol{x}_0$ 的预测区域两个问题.

(1) $\Theta\boldsymbol{z}_0$ 的 $1-\alpha$ 置信区域

由定理 2.4 知

$$\hat{\Theta} \sim N_{p,q+1}(\Theta, \Sigma, (ZZ')^{-1}), \quad Q \sim W_p(n - q - 1, \Sigma)$$

且相互独立. 所以

$$\hat{\Theta}\boldsymbol{z}_0 = XZ'(ZZ')^{-1}\boldsymbol{z}_0 \sim N_p(\Theta\boldsymbol{z}_0, \boldsymbol{z}_0'(ZZ')^{-1}\boldsymbol{z}_0 \cdot \Sigma)$$

且与 Q 相互独立. 于是由 Hotelling-T^2 统计量的定义知道

$$T^2 = (\boldsymbol{z}_0'(ZZ')^{-1}\boldsymbol{z}_0)^{-1}\boldsymbol{z}_0'(\hat{\Theta} - \Theta)'\left(\frac{Q}{n - q - 1}\right)^{-1}(\hat{\Theta} - \Theta) \cdot \boldsymbol{z}_0$$

$$\sim \frac{p(n - q - 1)}{n - q - p} F(p, n - q - p)$$

从而给定显著水平 α 后有

$$P\left\{T^2 \leqslant \frac{p(n-q-1)}{n-q-p} F_{1-\alpha}(p, n-q-p)\right\} = 1-\alpha$$

故 Θz_0 的 $1-\alpha$ 置信椭球为

$$(\hat{\Theta} z_0 - \Theta z_0)' Q^{-1} (\hat{\Theta} z_0 - \Theta z_0) \leqslant \frac{p}{n-q-p} \cdot z_0'(ZZ')^{-1} z_0 \cdot F_{1-\alpha}(p, n-q-p) \qquad (3.12)$$

根据第五章(2.5)式，对于 Θz_0 的各个分量，即对 $\theta^{(i)\prime} z_0$ $(i=1,2,\cdots,p)$，它们的 $1-\alpha$ 同时置信区间为

$$|\hat{\theta}^{(i)\prime} z_0 - \theta^{(i)\prime} z_0| \leqslant \sqrt{\frac{p}{n-q-p} F_{1-\alpha}(p, n-q-p) \cdot z_0'(ZZ')^{-1} z_0 q_{ii}} \qquad (3.13)$$

其中 $\hat{\theta}^{(i)\prime}$，$\theta^{(i)\prime}$ 分别为 $\hat{\Theta}$ 和 Θ 的第 i 行向量，q_{ii} 是 Q 的 (ii) 元素。

(2) x_0 的预测区域

由第二节关于预测问题的讨论知，在正态假设下，误差

$$x_0 - \hat{\Theta} z_0 \sim N_p[0, (1 + z_0'(ZZ')^{-1} z_0)\Sigma]$$

类似地有 x_0 的 $1-\alpha$ 预测区域是椭球

$$(x_0 - \hat{\Theta} z_0)' Q^{-1} (x_0 - \hat{\Theta} z_0) \leqslant (1 + z_0'(ZZ')^{-1} z_0) \cdot \frac{p}{n-p-q} F_{1-\alpha}(p, n-p-q) \qquad (3.14)$$

而 x_0 的分量的同时预测区间为

$$|x_{0i} - \hat{\theta}^{(i)\prime} z_0| \leqslant \sqrt{\frac{p}{n-p-q} F_{1-\alpha}(p, n-p-q) \cdot (1 + z_0'(ZZ')^{-1} z_0) q_{ii}} \qquad (3.15)$$

比较(3.12)与(3.14)，(3.13)与(3.15)，显然后者稍宽一些，它反映了随机误差 ε_0 的影响。

例 3.3　利用例 3.1 建立的回归方程来求出当 $z_0 = (1, 130,$

7.5)′ 时相应的预测值 $\hat{x}_0$，并给出显著水平 $\alpha=0.05$ 时 Θz_0 的 $1-\alpha$ 置信椭圆和 x_0 的 $1-\alpha$ 预测椭圆。

由回归方程

$$\begin{cases}\hat{x}_1=8.42+1.08z_1+0.42z_2\\ \hat{x}_2=14.42+2.25z_1+5.67z_2\end{cases}$$

将 $z_0=(1,130,7.5)'$ 代入得预测值为

$$\hat{x}_0=\begin{pmatrix}\hat{x}_{10}\\ \hat{x}_{20}\end{pmatrix}=\hat{\Theta}z_0=\begin{pmatrix}151.97\\ 349.45\end{pmatrix}$$

这里 $n=7$，$p=2$，$q+1=3$。由(3.12)知 Θz_0 的 0.95 置信椭圆为

$$(\mu_1-151.97,\ \mu_2-349.45)\begin{pmatrix}5.80 & 5.30\\ 5.30 & 13.13\end{pmatrix}^{-1}$$

$$\times\begin{pmatrix}\mu-151.97\\ \mu_2-349.45\end{pmatrix}\leqslant\frac{2}{3}\cdot 9.55$$

其中 $(\mu_1,\mu_2)'=(\theta^{(1)'}z_0,\theta^{(2)'}z_0)$，$0.347=z_0'(ZZ')^{-1}z_0$，$9.55=F_{0.95}(2,3)$。

对于 $x_0=(x_{10}\ \ x_{20})'$ 的 0.95 预测区域只须用

$$1+z_0'(ZZ')^{-1}z_0=1.347$$

代替上式中的 $z_0'(ZZ')^{-1}z_0=0.347$ 即可得。显然，Θz_0 的 0.95 置信椭圆与 x_0 的 0.95 预测椭圆其中心相同，但后者较大。

*§ 6.4 变量的筛选

在进行回归分析时，如果把某个对因变量有显著影响的自变量遗漏掉，回归分析的效果是不会好的，于是人们尽可能多的选择自变量，特别是在对研究的现象尚不清晰时往往会考虑许多自变量，如在气象学中考虑的自变量可能会超过 100 个。这样常常把对因变量没有显著影响的自变量列入了回归方程，实践证明，其回归效果不因为这些自变量的增加而变得更好，相反会带来许多麻

烦．因此在回归分析中自变量的选择是一个十分重要的问题，在本节里我们将介绍两种常用的变量筛选方法，即逐步回归方法和双重筛选的逐步回归方法．根据所介绍的计算步骤很容易编制相应的程序．

（一） 逐步回归方法

1. 逐步回归的基本思想

在建立回归方程时，我们总是希望得到的回归方程中包含了对因变量有显著影响的自变量，而不包含对因变量没有显著影响的自变量、不妨称这种回归方程为"最佳"回归方程，建立"最佳"回归方程我们自然会想到用下列两种方法：

（1）设可供选择的自变量有 q 个：$z_1, \cdots, z_q$，按 § 6.2 的方法先建立含有 q 个自变量的回归方程，再按 § 6.3 的方法对各个自变量进行显著性检验，把不重要的自变量剔除．例如第一次剔除一个自变量保留 $q-1$ 个自变量，从新建立含有 $q-1$ 个自变量的回归方程（注意原来的回归方程是在 q 个自变量的情况下计算的，对于 $q-1$ 个自变量的回归方程的回归系数必须重新计算），然后对含有 $q-1$ 个自变量的回归方程又进行检验，把作用不显著的那一个再剔除，再重新建立 $q-2$ 个自变量的回归方程……，终于有一个阶段没有再可以剔除的自变量了，这时便可得到最后的回归方程．

（2）先在 q 个自变量中选取一个对因变量最有影响的自变量 z_{i_1} 建立回归方程，然后再从余下的 $q-1$ 个自变量中选取另一个自变量 z_{i_2} 使得包含（z_{i_1} z_{i_2}）的回归方程效果最好，依次类推，直至 q 个自变量都被选入，这样得到了 q 个回归方程，然后按某种准则比较这 q 个回归方程，从中选择一个最好的．

上述两种方法都有缺点，首先是计算量非常大，另外，如果一个自变量被剔除后，它就永远被排斥在回归方程外，如果一个自变量被选入它就永远留在回归方程中，这样往往会产生不合理的结果．这是因为自变量之间也存在着某些制约关系，因此每个自变

量对因变量的影响的显著性不但与它自身有关，而且还与同它一起进入回归方程的其它自变量有关，例如某个 z_i 原先影响显著，但后来回归方程中的自变量发生变化牵制了 z_i 的影响，而使它变得不显著了(当然象这样的自变量还是得剔除)，相反地，若某个自变量 z_k 原先影响不显著，后来回归方程的自变量变化了，z_k 又变得显著了(象这样的自变量还得重新选入方程)。由上所述，在建立"最佳"回归方程的过程中通常有大量的自变量进进出出，一直到该进的进完了，该出的出完了为止。于是便产生了逐步回归的想法。

所谓逐步回归方法是指一种建立"最佳"回归方程的计算方法，它在计算的过程中同时考虑到自变量的选入和剔除，能以较少的计算量建立起"最佳"回归方程。此方法是从一个自变量开始，把自变量逐个选入，每一步只选一个自变量，并且要求当步选入的自变量是所有未被选入的自变量中最显著的一个。在选入自变量的同时还可以判断已在回归方程中的自变量是否还显著，若仍然显著，就让它继续保留在回归方程中，若变得不显著，那么立刻剔除。这样反复进行直到自变量既不能被选入又不能被剔除为止。并立即能建立起相应的"最佳"回归方程。

2. 选入和剔除自变量的检验

一般地，我们只须对某一步给出选入自变量与剔除自变量的方法，就可以得到一般的方法。在这一小段中我们先介绍选入和剔除自变量的依据。

假定在某一步已有 r 个自变量 $z_1, \cdots, z_r$ 入选，待考察是否能入选的自变量是 $z_{r+1}, \cdots, z_q$，相应的资料阵记为

$$X = \begin{pmatrix} x_{11} & x_{12} \cdots x_{1n} \\ x_{21} & x_{22} \cdots x_{2n} \\ \cdots & \cdots\cdots \\ x_{p1} & x_{p2} \cdots x_{pn} \end{pmatrix} = \begin{pmatrix} x^{(1)'} \\ x^{(2)'} \\ \vdots \\ x^{(p)'} \end{pmatrix}$$

$$Z=\begin{pmatrix} z_{11} & z_{12} \cdots z_{1n} \\ z_{21} & z_{22} \cdots z_{2n} \\ \cdots & \cdots \\ z_{r1} & z_{r2} \cdots z_{rn} \end{pmatrix}=\begin{pmatrix} z^{(1)'} \\ z^{(2)'} \\ \vdots \\ z^{(r)'} \end{pmatrix},$$

$$z^{(r+1)'}=(z_{(r+1)1}, z_{(r+1)2}, \cdots, z_{(r+1)n}), \cdots,$$

$$z^{(q)'}=(z_{q1}, z_{q2}, \cdots, z_{qn})$$

如果只考虑 $z_1, \cdots, z_r$ 对 $x_1, \cdots, x_p$ 的回归方程，就有

$$\begin{cases} X=(\theta_0 \Theta)\begin{pmatrix} 1_n' \\ Z \end{pmatrix}+\varepsilon \\ \varepsilon \sim N_{p,n}(0, \Sigma, I_n) \end{cases} \tag{4.1}$$

注意到(4.1)中 Θ 为对应于自变量 $z_1, \cdots, z_r$ 的回归系数矩阵，而 Z 是它们相应的资料矩阵。

逐个考察添加 $z_{r+1}, \cdots, z_q$ 就相当于把它们所对应的资料 $z^{(r+1)'}, \cdots, z^{(q)'}$ 添加在(4.1)式的右端，例如考察 $z^{(r+1)'}$ 添入后的模型，此时就成为

$$\begin{cases} X=(\theta_0 \ \Theta \ \theta_{r+1})\begin{pmatrix} 1_n' \\ Z \\ z^{(r+1)'} \end{pmatrix}+\varepsilon \\ \varepsilon \sim N_{p,n}(0, \Sigma, I_n) \end{cases} \tag{4.2}$$

显然，要问添加 $z^{(r+1)}$ 后对回归方程是否有显著影响，就得检验(4.2)中的 θ_{r+1} 是否等于 0，即检验 $H_0: \theta_{r+1}=0$ 是否成立。根据本章定理 3.2，检验统计量为

$$\Lambda_{r+1}=|Q_{r+1}|/|Q_r| \sim \Lambda(p, n-r-2, 1) \tag{4.3}$$

其中 Q_{r+1} 和 Q_r 分别为对应于(4.2)和(4.1)的回归方程的残差交叉乘积矩阵，再根据第三章表 3.1 知

$$F_{r+1}=\frac{n-r-p-1}{p} \cdot \frac{1-\Lambda_{r+1}}{\Lambda_{r+1}} \sim F(p, n-r-p-1)$$

这样检验 H_0: $\theta_{r+1}=0$ 就化成了 F 检验。对于给定的显著水平 α，由自由度 $(p, n-r-p-1)$ 查 F 表得临界值 $F_{1-\alpha}(p, n-$

$r-p-1$)，若 F_{r+1} 小于临界值，则认为 z_{r+1} 对因变量没有显著影响，否则有显著影响。

类似地，对 $z_{r+1},\cdots,z_q$ 中任一个 z_i 选入它以后是否有显著影响的检验统计量为

$$F_i=\frac{n-r-p-1}{p}\cdot\frac{1-\Lambda_i}{\Lambda_i}\sim F(p,n-r-p-1) \quad (4.4)$$

对于 $i=r+1,r+2,\cdots,q$ 计算出其 F 值：F_{r+1}，$F_{r+2},\cdots$，F_q。从中找出最大的一个，例如 F_s 最大，如果

$$F_s\leqslant F_{1-\alpha}(p,n-r-p-1) \quad (4.5)$$

则 z_s 不必引进。由于 F_s 是最大的一个，故其它自变量都没有引进的必要。若(4.5)不成立，则要引入 z_s，此时将(4.1)式中 Z 扩充一行 $z^{(s)\prime}$。再逐个考察余下的 $q-r-1$ 个自变量是否有可选入的，这一过程和刚才完全类似。

现在来讨论剔除自变量的检验。设已入选的自变量有 r 个：$z_1,z_2,\cdots,z_r$。完全类似于引入情况的讨论可得到已在回归方程中的第 j 个自变量的显著性检验统计量为

$$\tilde{F}_j=\frac{n-r-p}{p}\cdot\frac{1-\Lambda_j}{\Lambda_j}\sim F(p,n-r-p) \quad (4.6)$$

$$j=1,2,\cdots,r.$$

其中

$$\Lambda_j=|Q_r|/|Q_{r-1}|\sim\Lambda(p,n-r-1,1)$$

而 Q_r 与 Q_{r-1} 分别为回归方程中有自变量 z_j 和没有 z_j 的残差交叉乘积矩阵。

对于 $j=1,2,\cdots,r$ 计算出其 F 值：$\tilde{F}_1,\cdots,\tilde{F}_r$。对于给定的显著水平 α，根据自由度 $(p,n-r-p)$ 查 F 表得临界值 $F_{1-\alpha}(p,n-r-p)$，在 $\tilde{F}_1,\cdots,\tilde{F}_r$ 中找出最小的一个，例如 $\tilde{F}_t$ 最小，如果

$$\tilde{F}_t>F_{1-\alpha}(p,n-r-p) \quad (4.7)$$

则表明 z_t 对因变量有显著影响不应剔除。由于 $\tilde{F}_t$ 是最小的一个，

当然其它自变量就更显著了，若(4.7)不成立，则应将 z_i 从回归方程中剔除。剔除后再重复刚才的讨论，看看是否还有可剔除的，如果没有剔除的，则再转入"是否可引入新自变量"的讨论。这样反复进行直到出现既不能引入又不能剔除时逐步回归终止。

注意到上述"引入"自变量的检验中，z_i 还未被引入之前统计量(4.4) F_i 中的 Λ_i 事先是未知的，即(4.3)式中的 Q_{r+1} 是无法算出的，因此我们希望从已含有 r 个自变量的回归方程的残差交叉乘积矩阵 Q_r 导出 Q_{r+1}，关于它的推导可参看[9]。在具体实施时，它们之间的关系是通过一种紧凑的矩阵变换来实现(后面将要介绍)。此外还要注意到在"引入"与"剔除"的各步中都是用 F 检验，但其临界值是与已在方程中的自变量个数 r 有关，即每进行一步临界值都发生变化。但当 $n-r-p$ 较大时，$F_{1-\alpha}(p, n-r-p-1)$ 与 $F_{1-\alpha}(p,n-r-p)$ 相差不大，其中参数 r 为已在回归方程中的自变量个数，为了方便编制程序，根据经验往往把 r 固定的取为 $\frac{q}{2}$ 的整数值。这样无论是"引入"或者"剔除"，无论在那一步都采用同一个检验临界值

$$F^* = F_{1-\alpha}\left(p, n-p-\frac{q}{2}\right) \tag{4.8}$$

3. 逐步回归所用的紧凑变换法

逐步回归总是在对回归方程中所含自变量的基础上考虑引入新的自变量或剔除自变量，无论是引入或剔除都希望在施行后能很快地判断出下一步的动向，并且在逐步回归结束后能立刻写出最后的回归方程。下面介绍的紧凑变换法就能够做到这些要求。

设已取得 q 维自变量和 p 维因变量的数据矩阵，且已经中心化了(当然也可以考虑标准化数据，此时后面的(4.11)式便是相关系数矩阵)，把它们记为

$$Z = \begin{pmatrix} z_{11} & z_{12} \cdots z_{1n} \\ \cdots\cdots\cdots\cdots \\ z_{q1} & z_{q2} \cdots z_{qn} \end{pmatrix} = (z_1 \cdots z_n) \tag{4.9}$$

$$X = \begin{pmatrix} x_{11} & x_{12} \cdots x_{1n} \\ \cdots & \cdots \\ x_{p1} & x_{p2} \cdots x_{pn} \end{pmatrix} = (x_1 \cdots x_n) \qquad (4.10)$$

把 Z 和 X 合成一个 $(q+p)\times n$ 阶的资料矩阵

$$U = \begin{pmatrix} Z \\ X \end{pmatrix}$$

并计算 U 矩阵的交叉乘积矩阵

$$L = UU' = \begin{pmatrix} Z \\ X \end{pmatrix}(Z'X') = \begin{pmatrix} L_{ZZ} & L_{ZX} \\ L_{XZ} & L_{XX} \end{pmatrix}$$

$$= \left(\begin{array}{ccc|ccc} l_{11} & \cdots & l_{1q} & l_{1(q+1)} & \cdots & l_{1(q+p)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ l_{q1} & \cdots & l_{qq} & l_{q(q+1)} & \cdots & l_{q(q+p)} \\ \hline l_{(q+1)1} & \cdots & l_{(q+1)q} & l_{(q+1)(q+1)} & \cdots & l_{(q+1)(q+p)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ l_{(q+p)1} & \cdots & l_{(q+p)q} & l_{(q+p)(q+1)} & \cdots & l_{(q+p)(q+p)} \end{array}\right) \qquad (4.11)$$

其中 $L_{ZZ} = ZZ'$, $L_{XX} = XX'$, $L_{ZX} = ZX' = (XZ')' = L'_{XZ}$, 它们是 Z 和 X 的各种交叉乘积矩阵(若数据都是标准化了,则 L 便是 $z_1, \cdots, z_q$, $x_1, \cdots, x_p$ 的相关系数矩阵)。

紧凑变换法就是通过矩阵(4.11)来完成,即每一步,不论是引入或剔除一个自变量,都必须对(4.11)作一次矩阵变换,为了标明变换的步数,我们将(4.11)记为

$$L^{(0)} = (l_{ij}^{(0)})_{(q+p)\times(q+p)} \quad (i,j = 1,2,\cdots q,q+1,\cdots,q+p)$$

并用

$$L^{(m)} = (l_{ij}^{(m)})_{(q+p)\times(q+p)} \quad (i,j = 1,2,\cdots,q,q+1,\cdots,q+p)$$

表示对应于第 m 步变换后的相应矩阵。第 $m+1$ 步是在第 m 步的基础上引入或剔除自变量 z_r 而得来的(如果说第 $m+1$ 步是剔除变量 z_r, 那么说明在第 m 步以前回归方程已引入了 z_r; 如果说第 $m+1$ 步是引入自变量 z_r, 那么说明直到第 m 步 z_r 还在回归方程之外). 紧凑变换是指 $L^{(m+1)}$ 的元素 $l_{ij}^{(m+1)}$ 由 $L^{(m)}$ 的元素 $l_{ij}^{(m)}$ 按以下的变换公式而确定:

$$l_{ij}^{(m+1)}=\begin{cases}\dfrac{1}{l_{rr}^{(m)}} & \text{当 } i=j=r \text{ 时（}r\text{ 行 }r\text{ 列的交叉元素）}\\ -l_{ij}^{(m)}/l_{rr}^{(m)} & \text{当 } i\neq r,\ j=r \text{ 时（}r\text{ 列的元素）}\\ l_{ij}^{(m)}/l_{rr}^{(m)} & \text{当 } i=r,\ j\neq r \text{ 时（}r\text{ 行的元素）}\\ l_{ij}^{(m)}-l_{ir}^{(m)}l_{rj}^{(m)}/l_{rr}^{(m)} & \text{当 } i\neq r,\ j\neq r \text{ 时（其它行与列的元素）}\end{cases} \tag{4.12}$$

对矩阵 $L^{(m)}$ 按(4.12)进行变换，称它为对 $L^{(m)}$ 进行 r 消去变换，这个变换公式来源于解线性方程组的矩阵变换法．若记

$$L_{ix}^{(m+1)'}=(l_{i(q+1)}^{(m+1)}\cdots l_{i(q+p)}^{(m+1)}),\quad L_{xi}^{(m+1)}=(l_{(q+1)i}^{(m+1)}\cdots l_{(q+p)i}^{(m+1)})'$$

则矩阵变换公式(4.12)有下面的性质：

(i) 如果 z_i 是入选变量，则 $L_{ix}^{(m+1)}=-L_{xi}^{(m+1)}$。

(ii) 如果 z_i 是未入选变量，则 $L_{ix}^{(m+1)}=L_{xi}^{(m+1)}$。

(iii) $L_{XX}^{(m+1)}$ 是对应于 $m+1$ 步时的回归方程的残差交叉乘积矩阵，若 z_i 是入选的自变量，$L_{ix}^{(m+1)}$ 是相应于 z_i 的回归系数。

(iv) $m+1$ 步矩阵 $L^{(m+1)}$ 只与 $m+1$ 步自变量的全体有关，而与引入 $m+1$ 步自变量的先后顺序及曾经引入而后又被剔除的自变量无关，例如，到了第 $m+1$ 步已引入了 k 个自变量：$z_{i_1},\cdots,z_{i_k}$（称它们为 $m+1$ 步自变量的全体），那么 $L^{(m+1)}$ 只与这些自变量的全体有关，而与它们引入的先后顺序无关。

(v) 变换(4.12)对同一自变量接连施行两次消去变换，矩阵元素不变，即 $L^{(m+1)}=L^{(m-1)}$。

由性质(iv)和(v)可知，第 $m+1$ 步无论是引入或者剔除自变量 z_r，都要对前一步矩阵 $L^{(m)}$ 按(4.12)式进行 r 消去变换．当引入时，表明第 m 步时 z_r 不在回归方程中；当剔除时，表明第 m 时 z_r 已在回归方程中。

根据以上的性质，经过一些推算还可以得到如下两个结果：

(i) 如果在第 m 步已引入了 r 个自变量之后，要在下一步继续引入自变量的检验统计量(4.4)可写成

$$F_j^{(m+1)} = \frac{n-p-r-1}{p} \frac{L_{ix}^{(m)\prime}(L_{XX}^{(m)})^{-1}L_{xi}^{(m)}}{l_{ii}^{(m)} - L_{ix}^{(m)\prime}(L_{XX}^{(m)})^{-1}L_{xi}^{(m)}} \quad (4.13)$$

其中 i 为未进入回归方程的自变量足码。

(ii) 如果在第 m 步已有了 r 个自变量，下一步要剔除自变量的检验统计量(4.6)可写成

$$F_j^{(m+1)} = \frac{-(n-p-r)}{p} \frac{L_{ix}^{(m)\prime}(L_{XX}^{(m)\prime})L_{xj}^{(m)}}{l_{jj}^{(m)}} \quad (4.14)$$

其中 j 为进入回归方程的自变量足码。

4. 逐步回归具体计算步骤

准备工作：

由中心化数据矩阵(4.9)、(4.10)按照(4.11)计算

$$L^{(0)} = \begin{pmatrix} L_{ZZ}^{(0)} & L_{ZX}^{(0)} \\ L_{XZ}^{(0)} & L_{XX}^{(0)} \end{pmatrix} = (l_{ij}^{(0)})_{(q+p)\times(q+p)}$$

并记 $L_{ix}^{(0)\prime} = (l_{i(q+1)}^{(0)} \cdots l_{i(q+p)}^{(0)})$，对于给定的显著水平 α，根据(4.8)式确定“引入”和“剔除”自变量的检验临界值

$$F^* = F_{1-\alpha}\left(p, n-p-\frac{q}{2}\right)$$

第一步：引进第一个自变量。

按(4.13)式对所有的自变量 $z_i (i=1,2,\cdots,q)$ 计算

$$F_i^{(1)} = \frac{n-p-1}{p} \cdot \frac{L_{ix}^{(0)\prime}(L_{XX}^{(0)})^{-1}L_{xi}^{(0)}}{l_{ii}^{0} - L_{ix}^{(0)\prime}(L_{XX}^{(0)})^{-1}L_{xi}^{(0)}} \quad (i=1,2,\cdots,q)$$

找出一个最大的 $F_i^{(1)}$ 值来，若 $i=i_1$ 时 $F_{i_1}^{(1)}$ 最大，即

$$F_{i_1}^{(1)} = \max_i F_i^{(1)}$$

然后将 $F_{i_1}^{(1)}$ 与 F^* 比较，当 $F_{i_1}^{(1)} \leqslant F^*$ 时，停止筛选，转入结束，这种情况的出现说明自变量与因变量之间不存在线性关系；当 $F_{i_1}^{(1)} > F^*$ 时，引入 z_{i_1}。所谓引入 z_{i_1} 就是按公式(4.12)对矩阵 $L^{(0)}$ 进行 i_1 消去变换，并记成

$$L^{(0)} \xrightarrow{(i_1)} L^{(1)} \quad (4.15)$$

第二步：逐步引入变量。

假定已经入选了 r 个变量 $z_{i_1}, \cdots, z_{i_r}$，设这时矩阵 $L^{(0)}$ 已变成了 $L^{(m)}$。还有没有要入选的自变量呢？对未入选的自变量 $z_{i_{r+1}}, \cdots, z_{i_q}$，由 $L^{(m)}$ 按(4.13)式计算 F 值：$F_i^{(m+1)}$ $(i=i_{r+1}, \cdots, i_q)$，从中选出最大的 $F_i^{(m+1)}$ 来，若 $F_{i_{r+1}}^{(m+1)}$ 最大，再将它与 F^* 比较：

若 $F_{i_{r+1}}^{(m+1)} \leqslant F^*$，则停止引入转入第三步。

若 $F_{i_{r+1}}^{(m+1)} > F^*$，则继续引入变量 $z_{i_{r+1}}$，即按(4.12)对 $L^{(m)}$ 作 i_{r+1} 消去变换 $L^{(m)} \xrightarrow{(i_{r+1})} L^{(m+1)}$，由 $L^{(m+1)}$ 出发重复第二步。

第三步：逐步剔除变量

假定已入选了 $r+1$ 个自变量 $z_{i_1}, \cdots, z_{i_{r+1}}$，设这时 $L^{(0)}$ 已变成 $L^{(s)}$，考虑有没有需要剔除的自变量，对入选的自变量 $z_{i_1}, \cdots, z_{i_{r+1}}$，由 $L^{(s)}$ 按(4.14)式计算 F 值：$\tilde{F}_j^{(s+1)}$ $(j=i_1, i_2, \cdots, i_{r+1})$，从中选出一个最小的，若 $\tilde{F}_{i_t}^{(s+1)}$ 最小，将它与 F^* 比较：

若 $\tilde{F}_{i_t}^{(s+1)} > F^*$，表明不须要剔除 z_{i_t}，转入第二步。

若 $\tilde{F}_{i_t}^{(s+1)} \leqslant F^*$，表明要剔除 z_{i_t}，此时对 $L^{(s)}$ 按(4.12)式进行 i_t 消去变换 $L^{(s)} \xrightarrow{(i_t)} L^{(s+1)}$ 然后重复第三步。

这样反复进行，直到出现既不能引入又不能剔除时，逐步回归停止。

计算最后结果：

逐步回归结束后，必须写出最后得到的"最佳"回归方程、设最后选入了 m 个自变量 $z_{i_1}, z_{i_2}, \cdots, z_{i_m}$，这时 $L^{(0)}$ 已变成 $L^{(e)}$：

$$L^{(e)} = (l_{ij}^{(e)})_{(q+p)\times(q+p)}$$

根据矩阵变换公式(4.12)的性质（iii）立刻便可写出"最佳"回归方程为

$$\begin{cases} \hat{x}_1 = \bar{x}_1 + l_{i_1(q+1)}^{(e)}(z_{i_1} - \bar{z}_{i_1}) + \cdots + l_{i_m(q+1)}^{(e)}(z_{i_m} - \bar{z}_{i_m}) \\ \hat{x}_2 = \bar{x}_2 + l_{i_1(q+2)}^{(e)}(z_{i_1} - \bar{z}_{i_1}) + \cdots + l_{i_m(q+2)}^{(e)}(z_{i_m} - \bar{z}_{i_m}) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{x}_p = \bar{x}_p + l_{i_1(q+p)}^{(e)}(z_{i_1} - \bar{z}_{i_1}) + \cdots + l_{i_m(q+p)}^{(e)}(z_{i_m} - \bar{z}_{i_m}) \end{cases} \quad (4.15')$$

其中 $\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij},\ i = 1, 2, \cdots, p,\ \bar{x}_{i_1}, \cdots, \bar{x}_{i_m}$ 为自变量 $x_{i_1}, \cdots, x_{i_m}$ 的 n 个观察值的平均.

为了进行预测误差的分析，需要计算出"最佳"回归方程的残差交叉乘积矩阵 Q，仍然由性质 (iii) 便可得

$$Q = L_{XX}^{(l)} \tag{4.16}$$

注 当资料矩阵中的数据已是标准化数据，即 $L^{(0)}$ 为相关矩阵，那么最后的回归方程中各个变量都是标准化变量. 为了应用方便必须将它们还原为原变量(参看例 4.1).

我们就因变量的数目 $p = 1$ 的情况，举一个非常著名的例子，这个例子曾被许多书籍和文献引用. 注意到当 $p = 1$ 时(4.13)和(4.14)中的 L_{ix}, L_{XX}, L_{xi} 都变成了一个数.

例 4.1 某种水泥凝固时放出的热量 y (卡/克)与下列四种化

表 4.1 水泥试验数据

序号	1	2	3	4	5	6	7
x_1	7	1	11	11	7	11	3
x_2	26	29	56	31	52	55	71
x_3	6	15	8	8	6	9	17
x_4	60	52	20	47	33	22	6
y	78.5	74.3	104.3	87.6	95.9	109.2	102.7

序号	8	9	10	11	12	13
x_1	1	2	21	1	11	10
x_2	31	54	47	40	66	68
x_3	22	18	4	23	9	8
x_4	44	22	26	34	12	12
y	72.6	93.1	115.9	83.8	113.3	109.4

学成份有关：$z_1 = 3CaO \cdot Al_2O_3$（%），$z_2 = 3CaO \cdot SiO_2$（%），$z_3 = 4CaO \cdot Al_2O_3 \cdot Fe_2O_3$（%） $z_4 = 2CaO \cdot SiO_2$（%）。其试验数据见表 4.1 试用逐步回归方法建立回归方程（$\alpha = 0.1$）。

准备工作：

将表4.1的数据进行标准化处理，由标准化数据计算出 $L^{(0)}$（此时的 $L^{(0)}$ 就是 z_1, z_2, z_3, z_4, x 的相关系数矩阵）：

$$L^{(0)} = \begin{pmatrix} 1 & 0.228579 & -0.824134 & -0.245445 & 0.730717 \\ 0.228579 & 1 & -0.139242 & -0.972955 & 0.816253 \\ -0.824134 & -0.139242 & 1 & 0.029537 & -0.534671 \\ -0.245445 & -0.972955 & 0.029537 & 1 & -0.821305 \\ 0.730717 & 0.816253 & -0.534671 & -0.821305 & 1 \end{pmatrix}$$

根据所给的 $\alpha = 0.1$，$n = 13$，$p = 1$，$q = 4$，按(4.8)式查 F 表给出"引入"和"剔除"自变量的临界值为

$$F^* = F_{0.9}(1, 10) = 3.29$$

第一步：选择一个变量进入回归方程。利用 $L^{(0)}$ 及(4.13)式计算得

$$F_1^{(1)} = (13 - 2) \cdot \frac{l_{1x}^{(0)}(L_{XX}^{(0)})^{-1} l_{x1}^{(0)}}{l_{11}^0 - l_{1x}^{(0)}(L_{XX}^{(0)})^{-1} l_{x1}^{(0)}}$$

$$= 11 \times \frac{0.730717 \times 0.730717}{1 - 0.730717 \times 0.730717}$$

$$= 1.145620$$

同理可得

$$F_2^{(1)} = 21.960678, \quad F_3^{(1)} = 4.403423, \quad F_4^{(1)} = 22.7986$$

显然 $F_4^{(1)}$ 最大，且 $F_4^{(1)} > F^* = 3.29$，故选 z_4 进入回归方程。即对 $L^{(0)}$ 按(4.12)式进行 $i_1 = 4$ 的消去变换得

$$L^{(1)} = \begin{pmatrix} 0.939757 & -0.010228 & -0.816884 & 0.245445 & 0.529132 \\ -0.010228 & 0.053359 & -0.110504 & 0.972955 & 0.017160 \\ -0.816884 & -0.110504 & 0.999128 & -0.029537 & -0.510412 \\ -0.245445 & 0.972955 & 0.029537 & 1 & -0.821305 \\ 0.529132 & 0.017160 & -0.510412 & 0.821305 & 0.325458 \end{pmatrix}$$

第二步：继续逐个地引入自变量.

(1) 利用 $L^{(1)}$，按(4.13)式计算未进入回归方程的自变量 z_1, z_2, z_3 的 F 值：

$$F_1^{(2)} = (13-3)\frac{l_{1x}^{(1)}(L_{XX}^{(1)})^{-1}l_{x1}^{(1)}}{l_{11}^{(1)} - l_{1x}^{(1)}(L_{XX}^{(1)})^{-1}l_{x1}^{(1)}}$$

$$= 10 \times \frac{0.297930}{0.325459 - 0.297930} = 108.22$$

$$F_2^{(2)} = 0.172502,\quad F_3^{(2)} = 40.294854$$

显然 $F_1^{(2)}$ 最大，且 $F_1^{(2)} > F^*$，故引入 z_1，即对 $L^{(1)}$ 按(4.12)进行 $i_2 = 1$ 的消去变换得

$$L^{(2)} = \begin{pmatrix} 1.064105 & -0.010884 & -0.869250 & 0.261179 & 0.563052 \\ 0.010884 & 0.053284 & -0.119379 & 0.975626 & 0.022919 \\ 0.869250 & -0.119395 & 0.289051 & 0.183816 & -0.050463 \\ 0.261179 & -0.975626 & -0.183816 & 1.064105 & -0.683107 \\ -0.563052 & 0.050463 & -0.050463 & 0.683107 & 0.027529 \end{pmatrix}$$

(2) 继续引入自变量，利用 $L^{(2)}$，按(4.13)计算未进入回归方程的自变量 z_2, z_3 的 F 值：

$$F_2^{(3)} = 9 \times \frac{0.009865}{0.027529 - 0.009865} = 5.026324$$

$$F_3^{(3)} = 9 \times \frac{0.008810}{0.027529 - 0.008810} = 4.235803$$

显然 $F_2^{(3)}$ 最大，且 $F_2^{(3)} > F^*$。故再引入 z_2，即对 $L^{(2)}$ 按(4.12)进行 $i_3 = 2$ 的消去变换得

$$L^{(3)} = \begin{pmatrix} 1.66330 & 0.204391 & -0.893654 & 0.460589 & 0.567737 \\ 0.204391 & 18.780350 & -2.242271 & 18.322562 & 0.430415 \\ 0.893654 & 2.242271 & 0.021336 & 2.371429 & 0.000926 \\ 0.460589 & 18.322562 & -2.371430 & 18.940119 & -0.263182 \\ -0.567737 & -0.430415 & 0.000926 & 0.263182 & 0.017664 \end{pmatrix}$$

(3) 继续引入自变量，利用 $L^{(3)}$，按(4.13)计算 z_3 的 F 值：

$$F_3^{(4)} = 0.000390$$

显然 $F_3^{(4)} < F^*$，故 x_3 不能引入。

第三步：逐步剔除自变量。

由第二步知道已引入了三个自变量 x_4, x_1, x_2，由于回归方程中的自变量逐个增加，那么原来引进的自变量是否还显著？在这一步中要考虑剔除。

（1）利用 $L^{(3)}$ 及(4.14)式计算在回归方程中的自变量的 F 值：

$$\tilde{F}_4^{(4)} = -(13-3-1)\cdot\frac{l_{4x}^{(3)}(L_{XX}^{(3)})^{-1}l_{x4}^{(3)}}{l_{44}^{(3)}} = 1.863301$$

$$\tilde{F}_1^{(4)} = -(13-3-1)\cdot\frac{l_{1x}^{(3)}(L_{XX}^{(3)})^{-1}l_{x1}^{(3)}}{l_{11}^{(3)}} = 98.736394$$

$$\tilde{F}_2^{(4)} = -(13-3-1)\cdot\frac{l_{2x}^{(3)}(L_{XX}^{(3)})^{-1}l_{x2}^{(3)}}{l_{22}^{(3)}} = 5.026024$$

显然 $\tilde{F}_4^{(4)}$ 最小，且 $\tilde{F}_4^{(4)} < F^*$，故要剔除 x_4，即对 $L^{(3)}$ 按(4.12)进行 $i_1=4$ 的消去变换得

$$L^{(4)} = \begin{pmatrix} 1.055129 & -0.241181 & -0.835985 & -0.024318 & 0.574137 \\ -0.241181 & 1.055129 & 0.051847 & -0.967396 & 0.685017 \\ 0.835985 & -0.051847 & 0.318256 & -0.125207 & 0.033878 \\ 0.024318 & 0.967396 & -0.125207 & 0.052798 & -0.013896 \\ -0.574137 & -0.685017 & 0.033878 & -0.013896 & 0.021322 \end{pmatrix}$$

（2）继续剔除自变量，利用 $L^{(4)}$ 及(4.14)计算 x_1，x_2 的 F 值：

$$\tilde{F}_1^{(5)} = 146.520,\quad \tilde{F}_2^{(5)} = 208.578$$

显然最小者 $\tilde{F}_1^{(5)}$ 都大于 F^*，故不能再剔除。

第四步：在 $L^{(4)}$ 的基础上重复第二步，即考虑未进入回归方程的 x_3, x_4 是否还能引入。按(4.13)式计算得

$$F_3^{(5)} = 1.586729,\quad F_4^{(5)} = 1.61279.$$

它们都小于 F^*，故不能再引入。到此为止已出现既不能剔除又不能引入的情况，逐步回归停止。

给出最后结果：

由 $L^{(4)}$ 及(4.15),(4.16)式(注意此例的 $p=1$) 知道回归系数及相应的残差平方和为

$$\hat{\theta}_1^* = l_{13}^{(4)} = 0.574137, \quad \hat{\theta}_2^* = l_{23}^{(4)} = 0.685017,$$

$$Q^* = l_{33}^{(4)} = 0.021322$$

由于本例的变量是经过标准化,为了方便应用,我们要将它还原为原变量,为此记

$$S_{xx} = \sum_{j=1}^{n}(x_j - \bar{x})^2, \quad S_{ii} = \sum_{j=1}^{n}(z_{ij} - \bar{z}_i)^2, \quad i = 1,2.$$

容易证明还原后的回归系数 $\hat{\theta}_i$ 及残差平方和 Q 与标准化后的回归系数 $\hat{\theta}_i^*$ 及残差平方和 Q^* 有如下关系

$$\hat{\theta}_i = \sqrt{\frac{S_{xx}}{S_{ii}}}\,\hat{\theta}_i^* \tag{4.17}$$

$$Q = S_{xx}Q^* \tag{4.18}$$

由表(4.1)的原始数据可算得

$$S_{xx} = 2715.7635, \quad S_{11} = 415.2308, \quad S_{22} = 2905.6923$$

所以

$$\hat{\theta}_1 = \sqrt{\frac{2715.7635}{415.2308}} \times 0.574137 = 1.4683$$

$$\hat{\theta}_2 = \sqrt{\frac{2715.7635}{2905.6923}} \times 0.685017 = 0.6623$$

$$Q = 2715.7635 \times 0.021322 = 57.9055$$

再由(2.40)式可得

$$\hat{\theta}_0 = \bar{x} - \hat{\theta}_1\bar{z}_1 - \hat{\theta}_2\bar{z}_2 = 52.5773$$

从而得到最后的回归方程为

$$\hat{x} = 52.5773 + 1.4683z_1 + 0.6623z_2 \tag{4.19}$$

其残差的标准差为

$$\hat{\sigma} = \sqrt{\frac{Q}{13-2-1}} = 2.406$$

如果不剔除 z_3, z_4, 而按照 4 个自变量直接建立的回归方程

为

$$\tilde{x} = 62.4052 + 1.5511z_1 + 0.5101z_2 + 0.1019z_3 - 0.1441z_4 \tag{4.20}$$

其残差平方和及残差标准差为

$$\tilde{Q} = 47.86,\ \tilde{\sigma} = \sqrt{\frac{\tilde{Q}}{13-4-1}} = 2.446$$

比较两个回归方程的残差标准差，它们是很接近的。这说明其回归效果相差不多，但逐步回归得到的回归方程(4.19)只有两个自变量,应用起来方便多了。此外,我们对水泥放热量与其化学成份之间的关系也有了更进一步的了解。

***(二） 双重筛选逐步回归**

1. 问题的提出

在（一）中所讨论的逐步回归方法已被实际工作者广泛地应用,但这种方法有一个明显的不足,我们考察自变量 $z_1,\cdots,z_q$ 对因变量 $x_1,\cdots,x_p$ 的回归时，如果某个自变量 z_i 对一个因变量 x_j 有显著影响，对其它因变量 $x_1,\cdots,x_{j-1},x_{j+1},\cdots,x_p$ 没有显著影响,那么在逐步回归过程中作 F 检验时,很可能不显著而最终被淘汰掉。一般情况可以是这样的：因变量的一部分与自变量的一部分有密切的关系,例如 $x_1,\cdots,x_{p_1}$ 与 $z_1,\cdots,z_{q_1}$ 有密切关系，而另一部分因变量 $x_{p_1+1},\cdots,x_{p_2}$ 与 $z_{q_1+1},\cdots,z_{q_2}$ 有较密切的关系,…,如此等等。要注意的是 $x_1,\cdots,x_{p_1}$ 与 $x_{p_1+1},\cdots,x_{p_2}$ 不会有共同的变量,但 $z_1,\cdots,z_{q_1}$ 与 $z_{q_1+1},\cdots,z_{q_2}$，可能有共同的变量．因为某一个自变量可能对许多不同的因变量甚至全部因变量都有影响.因此自然地提出一个问题,能否有一种逐步回归的算法,将因变量分组,并找出对每组因变量有影响的自变量，然后对各组因变量与其相应有影响的自变量建立回归方程式。双重筛选逐步回归就是解决上述问题的一种方法。关于上述问题所对应的模型的描述可参看[9],在这里仅给出双重筛选逐步回归的计算步骤及一个实例。

2. 计算步骤

设我们已有 q 个自变量和 p 个因变量的中心化数据（当然也可以采用标准化数据，此时(4.11)式便是相关矩阵）矩阵(4.9)和(4.10)，并将它们合成一个 $(q+p)\times n$ 阶的资料矩阵

$$U=\begin{pmatrix} Z \\ X \end{pmatrix}$$

准备工作：

由中心化资料矩阵 U 按(4.11)式计算得

$$L^{(0)}=\begin{pmatrix} L_{ZZ}^{(0)} & L_{ZX}^{(0)} \\ L_{XZ}^{(0)} & L_{XX}^{(0)} \end{pmatrix}=(l_{ij}^{(0)})_{(q+p)\times(q+p)}$$

我们需要两个矩阵 L_1 与 L_2 来分别记录自变量 $z_1,\cdots,z_q$ 和因变量 $x_1,\cdots,x_p$ 的筛选情况。以下注意，我们是按合并资料矩阵 U 中自变量和因变量的顺序进行统一编号，例如第 i 号 $(i\leqslant q)$ 变量表示自变量 z_i，第 $q+j$ $(j\leqslant p)$ 号变量表示因变量 x_j。矩阵 $L^{(m)}$ 对某一编号 $k(k\leqslant q+p)$ 按公式(4.12)进行消去变换而得到 $L^{(m+1)}$，仍记为

$$L^{(m)}\xrightarrow{(k)}L^{(m+1)}$$

在双重筛选中还要涉及到只对 L 中的部分子矩阵进行消去变换，故用记号

$$L^{(m)}\xrightarrow{(k,*)}L^{(m+1)}$$

表示对 $L^{(m)}$ 中的三块子矩阵 $L_{ZZ}^{(m)}$，$L_{ZX}^{(m)}$，$L_{XZ}^{(m)}$ 进行 k 消去变换，而子块 $L_{XX}^{(m)}$ 保持不变。类似地用记号

$$L^{(m)}\xrightarrow{(k,X)}L^{(m+1)}$$

表示只对 $L^{(m)}$ 中的子块 $L_{XX}^{(m)}$ 进行 k 消去变换，而其它三块 $L_{ZZ}^{(m)}$、$L_{ZX}^{(m)}$、$L_{XZ}^{(m)}$ 保持不变。

因为在计算过程要对自变量和因变量进行筛选，故取两个临界值（注意这两个临界值随着变量个数的变化而不同，但为了方便，根据经验取成两个常数）：

$$F_x^* = F_{1-\alpha}\left(\frac{p}{2}, n - \frac{q+p}{2}\right),$$

$$F_x^* = F_{1-\beta}\left(\frac{q}{2}, n - \frac{(q+p)}{2}\right) \tag{4.21}$$

其中α和β为给定的显著水平。并且令

$$L_1^{(0)} = L_2^{(0)} = L^{(0)} \tag{4.22}$$

每建立了一组因变量的回归方程，就称为一个阶段的双重筛选结束。

第一步：任意选入一个因变量x_j，即对$L_1^{(0)}$和$L_2^{(0)}$进行下面的消去变换

$$L_1^{(0)} \xrightarrow{(q+j,X)} L_1^{(1)}, \quad L_2^{(0)} \xrightarrow{(q+j)} L_2^{(1)} \tag{4.23}$$

然后转入第二步。

第二步：逐个检验是否要引入自变量。

设这时已引入了p_0个因变量，q_0个自变量，$L_1^{(0)}$与$L_2^{(0)}$已相应地变成$L_1^{(m)}$和$L_2^{(m)}$，计算

$$A_i = 1 - \frac{l_{2ii}^{(m)}}{l_{1ii}^{(m)}} \quad (i = 1,2,\cdots,q) \tag{4.24}$$

其中$l_{1ii}^{(m)}, l_{2ii}^{(m)}$分别为$L_1^{(m)}$和$L_2^{(m)}$的$(i,i)$元素。注意到对已入选的自变量的$A_i < 0$，对未入选的自变量的$A_i \geqslant 0$，在大于零的$A_i$中找出最大的，设为$A_{i_1}$，计算

$$F_{i_1} = \frac{n - p_0 - q_0 - 1}{p_0} \cdot \frac{A_{i_1}}{1 - A_{i_1}}$$

如果$F_{i_1} \leqslant F_x^*$，不引入x_{i_1}，再转入第四步。

如果$F_{i_1} > F_x^*$，则引入x_{i_1}，即做下述消去变换

$$L_1^{(m)} \xrightarrow{(i_1,Z)} L_1^{(m+1)}; \quad L_2^{(m)} \xrightarrow{(i_1)} L_2^{(m+1)} \tag{4.25}$$

然后转入第三步。

第三步：逐个检验是否要剔除自变量。

设这时已引入了p_0个因变量，q_0个自变量，$L_1^{(0)}$与$L_2^{(0)}$已

相应地变成 $L_1^{(m)}$，$L_2^{(m)}$. 计算

$$\Lambda_i = 1 - l_{2ii}^{(m)}/l_{1ii}^{(m)} \quad (i = 1,2,\cdots,q) \tag{4.26}$$

对已入选的自变量的 $\Lambda_i < 0$，未入选的自变量 $\Lambda_i \geqslant 0$，在小于零的 Λ_i 中找最大的(绝对值最小)，记为 Λ_{i_1}，并计算

$$\tilde{F}_{i_1} = \frac{-(n - p_0 - q_0)}{p_0} \Lambda_{i_1} \tag{4.27}$$

若 $\tilde{F}_{i_1} \leqslant F_z^*$，则剔除 z_{i_1}，即做消去变换

$$L_1^{(m)} \xrightarrow{(i_1 Z)} L_1^{(m+1)}, \quad L_2^{(m)} \xrightarrow{(i_1)} L_2^{(m+1)}$$

然后转入第二步.

若 $\tilde{F}_{i_1} > F_z^*$，则不剔除 z_{i_1}. 转入第二步.

第四步：逐个检验是否要剔除因变量.

设这时已经引入了 p_0 个因变量，q_0 个自变量，$L_1^{(0)}$ 与 $L_2^{(0)}$ 已变成 $L_1^{(m)}$ 和 $L_2^{(m)}$，计算

$$\Lambda_{q+j} = 1 - l_{2(q+j)(q+j)}^{(m)}/l_{1(q+j)(q+j)}^{(m)} \quad (j \leqslant p) \tag{4.28}$$

对于已选入的因变量 x_j(即 $q + j$ 变量)的 $\Lambda_{q+j} < 0$，对未选入的因变量的 $\Lambda_{q+j} \geqslant 0$. 在小于零的 Λ_{q+j} 中找一个最大的绝对值最小)，设为 $q + j_1$，并计算

$$\tilde{F}_{q+j_1} = -\frac{(n - p_0 - q_0)}{p_0} \Lambda_{q+j_1} \tag{4.29}$$

若 $\tilde{F}_{q+j_1} \leqslant F_x^*$，则剔除 x_{j_1}，即做消去变换

$$L_1^{(m)} \xrightarrow{(q + j_1, X)} L_1^{(m+1)}, \quad L_2^{(m)} \xrightarrow{(q + j_1)} L_2^{(m+1)} \tag{4.30}$$

然后转入第三步.

若 $\tilde{F}_{q+j_1} > F_x^*$，则不剔除 x_{j_1}，转入第五步.

第五步：逐个检验是否要引入因变量.

设这时已引入 p_0 个因变量，q_0 个自变量，$L_1^{(0)}$ 和 $L_2^{(0)}$ 相应已变成 $L_1^{(m)}$ 和 $L_2^{(m)}$，计算

$$\Lambda_{q+j} = 1 - l_{2(q+j)(q+j)}^{(m)}/l_{1(q+j)(q+j)}^{(m)} \quad (j \leqslant p) \tag{4.31}$$

同样，对已入选的因变量都有 $\Lambda_{q+j} < 0$，而对未入选的因变量都

有 $\Lambda_{q+j} \geqslant 0$，在大于零的 Λ_{q+j} 找出最大的，记为 Λ_{q+j_1} 并计算

$$F_{q+j_1} = \frac{n - p_0 - q_0 - 1}{q_0} \frac{\Lambda_{q+j_1}}{1 - \Lambda_{q+j_1}}$$

若 $F_{q+j_1} \leqslant F_x^*$，则不引入 x_{j_1}，转入第六步。

若 $F_{q+j_1} > F_x^*$，则引入 x_{j_1}，即做消去变换

$$L_1^{(m)} \xrightarrow{(q+j_1,X)} L_1^{(m+1)}, \quad L_2^{(m)} \xrightarrow{(q+j_1)} L_2^{(m+1)}$$

然后转入第三步。'

第六步：给出第一组因变量的回归方程式。

当出现既不能剔除又不能引入变量时，第一阶段的双重筛选停止。此时 $L_1^{(0)}$ 与 $L_2^{(0)}$ 变成 $L_1^{(e)}$ 和 $L_2^{(e)}$，设选入的因变量为 $x_{j_1},\cdots,x_{j_k}$，选入的自变量为 $z_{i_1},\cdots,z_{i_t}$，则由 $L_1^{(e)}$ 可写出第一组回归方程式

$$\begin{cases} \hat{x}_{j_1} = \bar{x}_{j_1} + l_{1i_1(q+j_1)}^{(e)}(z_{i_1} - \bar{z}_{i_1}) + \cdots + l_{1i_t(q+j_1)}^{(e)}(z_{i_t} - \bar{z}_{i_t}) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{x}_{j_k} = \bar{x}_{j_k} + l_{1i_1(q+j_k)}^{(e)}(z_{i_1} - \bar{z}_{i_1}) + \cdots + l_{1i_t(q+j_k)}^{(e)}(z_{i_t} - \bar{z}_{i_t}) \end{cases} \tag{4.32}$$

第一阶段的双重筛选结束后，去掉第一阶段入选的因变量 $x_{j_1},\cdots,x_{j_k}$，用剩下的因变量和全部自变量重复上述的各步骤进行第二阶段的筛选，建立起第二组回归方程式，…如此类推，直到全部因变量用完，这样就结束了整个双重筛选逐步回归。

例 4.2　为了分析自然经济及气候条件对森林覆盖面积的消长影响，取了 12 个点进行调查，共测定了 12 变量，其数据列于表 4.2，表中变量的意义是：

z_1 — 山地比例(%)，z_2 — 人口密度（人/km²），z_3 — 人均收入增长率(元/年)，z_4 — 公路密度（百米/公顷），z_5 — 前汛期降水量(厘米/年)，z_6 — 后汛期降水量（厘米/年），z_7 — 月平均最低温度。

x_1 — 森林覆盖率(%)，x_2 — 针叶林比例(%)，x_3 — 造林面积(千亩/年)，x_4 — 年采伐面积(千亩/年)，x_5 — 火灾频数（次/

表 4.2　例 4.2 中各变量的数据表

编号	z_1	z_2	z_3	z_4	z_5	z_6	z_7	x_1	x_2	x_3	x_4	x_5
1	74.3	91.0	5.76	1.3	108	66	17.4	51.2	9.5	15.39	12.6	1
2	70.4	157.0	8.04	2.2	126	68	17.2	52.5	24.2	10.84	8.4	0
3	78.7	77.0	7.94	2.0	114	63	17.0	62.9	22.8	13.824	9.8	0
4	78.9	67.0	6.86	1.5	110	55	17.0	64.3	25.1	34.567	14.0	3
5	49.1	91.0	4.92	1.5	92	49	16.5	39.3	10.7	7.405	5.6	2
6	57.6	219.0	5.56	2.5	91	48	16.8	37.3	37.3	9.118	2.8	0
7	53.1	221.0	7.42	3.9	90	45	16.8	30.0	27.0	8.638	2.8	4
8	70.1	123.0	5.38	3.1	123	59	17.0	47.8	34.6	81.64	11.2	5
9	86.6	45.0	12.54	1.2	105	57	14.8	69.0	37.3	23.95	11.2	0
10	82.2	81.0	13.24	1.6	131	61	15.9	62.3	16.5	33.598	16.8	0
11	76.8	90.0	10.70	1.5	131	69	15.8	67.6	22.2	8.93	9.8	0
12	88.9	63.0	1.98	1.8	107	65	14.5	79.3	42.1	58.968	3.5	0

年).

按逐步回归程序,取 $F_1^* = 2$, $F_2^* = 2.5$, 这时将 5 个因变量分成了三组. 其相应的回归方程为

第一组:

$$\begin{cases} \hat{x}_1 = 52.91 + 0.820z_1 - 0.057z_2 - 0.425z_3 - 2.883z_7 \\ \hat{x}_4 = -47.79 + 0.130z_1 - 0.041z_2 + 0.623z_3 + 2.890z_7 \end{cases}$$

第二组:

$$\begin{cases} \hat{x}_2 = 72.68 + 0.435z_1 + 0.084z_2 + 5.444z_4 - 6.017z_7 \\ \hat{x}_5 = -2.05 - 0.067z_1 - 0.036z_2 + 2.659z_4 + 0.416z_7 \end{cases}$$

第三组:

$$\hat{x}_3 = -38.99 + 0.928z_1 - 0.296z_2 - 2.865z_3 + 26.026z_4$$

由上述计算结果表明,森林覆盖率 (x_1) 和采伐面积 (x_4) 受

到相同自变量的影响，即受到山地比例（z_1），人口密度（z_2），人均收入增长率（z_3）及月平均最低温度（z_7）的影响；针叶林比例（x_2）和火灾频数（x_5）主要受到公路密度（z_4）和最低气温（z_7）影响；而造林面积（x_3）主要受到公路密度（z_4）和人均收入增长率（z_3）的影响。

§6.5 多元方差分析

在一元统计中，我们考虑一个因素对试验结果(单个指标)是否有影响，这就是单向分类问题；若考虑两个因素对试验结果（单个指标)是否有影响，这就是双向分类问题，它们所用的检验统计量都是F统计量。如果上述的试验结果是多指标的，就是多元方差分析的任务。它的思想方法与一元完全相似，只是统计量改为 Wilks 的Λ统计量、在本节里我们将对多元方差分析中的单向分类问题进行较详细的讨论，而对双向分类问题作方法上的介绍，不作详尽的证明。

（一） 单向分类

在讨论单向多元方差分析之前，我们回顾一下单向一元方差分析是十分有益的。

考虑因素A对试验结果x是否有显著影响，设A取m个水平(等级)：$A_1,\cdots,A_m$，在第i个水平A_i之下做n_i次独立试验，用$x_{ij}(i=1,2,\cdots,m;\ j=1,2,\cdots,n_i)$表示在$A_i$之下的第$j$次试验结果。每水平之下视为一个总体，而该水平之下所做的n_i次试验为该总体的一个容量为n_i的样本。设这m个总体都服从正态分布$N(\mu_i,\sigma^2)$ $(i=1,2,\cdots,m)$。要推断A对试验结果有无显著影响，就等价于检验：

$$H_0:\mu_1=\mu_2=\cdots=\mu_m,\ H_1:\mu_i \text{ 不全相等} \qquad (5.1)$$

为了方便分析，引入一个量$\tau_i=\mu_i-\mu$，这里

$$\mu=\frac{1}{n}\sum_{i=1}^{m}n_i\mu_i,\quad n=\sum_{i=1}^{m}n_i.$$

这样 $\mu_i = \mu + \tau_i$，并称 τ_i 为因素 A 的第 i 个水平的效应。此时，检验(5.1)就等价于检验

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_m = 0, \ H_1: \tau_i \text{ 不全为零} \tag{5.2}$$

如何来检验(5.2)呢？必须先给出模型，由前面所述第 i 个总体的第 j 个观察值具有下述形式

$$\begin{cases} x_{ij} = \mu + \tau_i + \varepsilon_{ij} \\ \varepsilon_{ij} \sim N(0, \sigma^2) \ (\text{或 } x_{ij} \sim N(\mu + \tau_i, \sigma^2)) \end{cases} \tag{5.3}$$

称(5.3)为单向一元方差分析模型，其中 ε_{ij} 为误差。对于单向一元方差分析正是基于(5.3)对于具体试验数据（仍记为 x_{ij}）进行分解：

$$x_{ij} = \bar{x}_{..} + (\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.}) \tag{5.4}$$

其中 $\bar{x}_{..}$ 为全部观察数据的平均，$\bar{x}_{i.}$ 为第 i 个总体的 n_i 个观察数据的平均。(5.4)式右端三项分别作为(5.3)中的 μ, τ_i 和 ε_{ij} 的估计。(5.4)式也可以写成

$$x_{ij} - \bar{x}_{..} = (\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.}) \tag{5.5}$$

对于(5.5)式可以这样来解释：左边 $(x_{ij} - \bar{x}_{..})$ 是数据 x_{ij} 对 $\bar{x}_{..}$ 的偏离，这个偏离可分解为两部分组成，一部分为 A_i 水平效应引起的，即 $(\bar{x}_{i.} - \bar{x}_{..})$，一部分是误差引起的，即 $(x_{ij} - \bar{x}_{i.})$。我们将(5.5)两边平方后求和仍有同样的结构式，事实上

$$\begin{aligned} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2 &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} [(\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.})]^2 \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} (\bar{x}_{i.} - \bar{x}_{..})^2 + 2 \sum_{i=1}^{m} \sum_{j=1}^{n_i} (\bar{x}_{i.} - \bar{x}_{..})(x_{ij} - \bar{x}_{i.}) \\ &\quad + \sum_{i=1}^{m} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} (\bar{x}_{i} - \bar{x}_{..})^2 + \sum_{i=1}^{m} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 \end{aligned} \tag{5.6}$$

上式最后一个等号是因为交叉乘积求和项为 0。记

$$S=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{..})^2,\quad S_A=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\bar{x}_{i.}-\bar{x}_{..})^2,$$

$$S_e=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2$$

并分别称为总的离差平方和，效应平方和以及误差平方和。从而(5.6)式可简写为

$$S=S_A+S_e \tag{5.7}$$

要检验假设(5.2)只须把 S_A 与 S_e 进行比较。检验统计量为

$$F=\frac{S_A/m-1}{S_e/n-m} \tag{5.8}$$

在一元方差分析中已证明(5.8)式的 F 当 H_0 成立时是服从自由度为 $(m-1,\ n-m)$ 的 F 分布，因此对于临界值 $F_{1-\alpha}(m-1,n-m)$。若

$$F>F_{1-\alpha}(m-1,n-m)$$

则拒绝 H_0，即因素 A 对试验结果有显著影响，否则没有显著影响。通常我们把上述过程列成一个方差分析表(表 5.1)。

表 5.1　单向一元方差分析表

方差来源	平　方　和	自由度	平均方差	F 值
因素的影响	$S_A=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\bar{x}_{i.}-\bar{x}_{..})^2$	$m-1$	$S_A/m-1$	$\frac{S_A/m-1}{S_e/n-m}$
误　差	$S_e=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2$	$n-m$	$S_e/n-m$	
总　和	$S=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{..})^2$	$n-1$		

与一元情况相类似，我们考虑单向的多元方差分析，设试验结果是多指标，即视为一个向量，不妨设为 p 维，因素 A 取 m 个水平

$A_1, \cdots, A_m$. 在第 i 个水平下作 n_i 次独立试验，用向量 $\boldsymbol{x}_{ij}$ 表示在第 i 水平 A_i 下做的第 j 次试验结果，把同一水平下的 n_i 个试验结果视为同一总体的一个容量为 n_i 的样本。设这 m 个总体都是 p 维正态总体 $N_p(\boldsymbol{\mu}_i, \Sigma)$ $(i=1,2,\cdots,m)$，它们有公共的协方差矩阵 Σ。要推断 A 对试验结果有无影响等价于检验

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_m,\quad H_1: \boldsymbol{\mu}_i \text{ 不全为 } 0 \tag{5.9}$$

引入因素水平效应向量 $\boldsymbol{\tau}_i = \boldsymbol{\mu}_i - \boldsymbol{\mu}$ $\left(\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{m} n_i \boldsymbol{\mu}_i\right)$后，(5.9)可写成

$$H_0: \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_m = 0,\quad H_1: \boldsymbol{\tau}_i \text{ 不全为 } 0 \tag{5.10}$$

由上所述第 i 个总体的第 j 个观察向量具有下列形式

$$\begin{cases} \boldsymbol{x}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\varepsilon}_{ij} \\ \boldsymbol{\varepsilon}_{ij} \sim N_p(0, \Sigma) \ (\text{或 } \boldsymbol{x}_{ij} \sim N_p(\boldsymbol{\mu} + \boldsymbol{\tau}_i, \Sigma)) \end{cases} \tag{5.11}$$

其中 $\boldsymbol{\varepsilon}_{ij}$ 为误差向量。称(5.11)为单向多元方差分析模型，显然是线性模型。注意到 $\sum_{i=1}^{m} n_i \boldsymbol{\tau}_i = \sum_{i=1}^{n} n_i \boldsymbol{\mu}_i - n\boldsymbol{\mu} = 0$，并记

$$\bar{\boldsymbol{x}}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij},\quad \bar{\boldsymbol{x}}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} = \frac{1}{n}\sum_{i=1}^{m} n_i \bar{\boldsymbol{x}}_{i\cdot}$$

根据模型(5.11)，对于具体的观察向量(仍记为 $\boldsymbol{x}_{ij}$) 可分解为

$$\boldsymbol{x}_{ij} = \bar{\boldsymbol{x}}_{\cdot\cdot} + (\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot}) + (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i\cdot})$$

或

$$\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{\cdot\cdot} = (\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot}) + (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i\cdot}) \tag{5.12}$$

这个分解导致了交叉乘积 $(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{\cdot\cdot})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{\cdot\cdot})'$ 可分解为

$$\begin{aligned}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{\cdot\cdot})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{\cdot\cdot})' &= [(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot}) \\ &\quad + (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i\cdot})][(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot}) + (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i\cdot})]' \\ &= (\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot})(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot})' + (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i\cdot})(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot})' \\ &\quad + (\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i\cdot})' + (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i\cdot})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i\cdot})'\end{aligned}$$

注意到

$$\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{i\cdot})=0,\quad \sum_{i=1}^{m}n_i(\bar{\boldsymbol{x}}_{i\cdot}-\bar{\boldsymbol{x}}_{\cdot\cdot})=0 \tag{5.13}$$

然后将上式对 i,j 求和便得到

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{\cdot\cdot})(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{\cdot\cdot})'$$

$$=\sum_{i=1}^{m}n_i(\bar{\boldsymbol{x}}_{i\cdot}-\bar{\boldsymbol{x}}_{\cdot\cdot})(\bar{\boldsymbol{x}}_{i\cdot}-\bar{\boldsymbol{x}}_{\cdot\cdot})'$$

$$+\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{i\cdot})(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{i\cdot})' \tag{5.14}$$

以 Q,B,E_i 和 E 分别记

$$Q=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{\cdot\cdot})(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{\cdot\cdot})',$$

$$B=\sum_{i=1}^{m}n_i(\bar{\boldsymbol{x}}_{i\cdot}-\bar{\boldsymbol{x}}_{\cdot\cdot})(\bar{\boldsymbol{x}}_{i\cdot}-\bar{\boldsymbol{x}}_{\cdot\cdot})'$$

$$E_i=\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{i\cdot})(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{i\cdot})',$$

$$E=\sum_{i=1}^{m}E_i=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{i\cdot})(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{i\cdot})'$$

则(5.14)可简写为

$$Q=B+E \tag{5.15}$$

仿照一元情况的类似称呼，称 Q 为总离差"平方和"矩阵，B 为因素的效应"平方和"矩阵，E 为误差"平方和"矩阵。

为了今后讨论的方便，令

$$C=I_n-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n',\quad C_i=I_{n_i}-\frac{1}{n_i}\mathbf{1}_{n_i}\mathbf{1}_{n_i}',\quad M_i=\boldsymbol{\mu}_i\mathbf{1}_{n_i}',$$

$$M=(M_1,\cdots,M_m),\quad M_*=\boldsymbol{\mu}\mathbf{1}_n',\quad X=(X_1,\cdots,X_m)$$

其中 $X_i=(\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{in_i})$ 为第 i 个总体的样本矩阵．那么上面记号中的 X 就是 m 个总体的样本组成的 $p\times n$ 阶矩阵．这样，模

型(5.11)可表示为

$$X \sim N_{p,n}(M, \Sigma, I_n) \quad (\Sigma > 0) \tag{5.16}$$

检验(5.9)(或者(5.10))就等价检验

$$H_0:\ M = M_0 \qquad H_1:\ M \neq M_0 \tag{5.17}$$

注意到引入上述记号后，还有下面的关系式成立:

$$Q = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})(x_{ij} - \bar{x}_{..})' = XCX \tag{5.18}$$

$$E_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})(x_{ij} - \bar{x}_{i.})' = X_i C_i X_i' \tag{5.19}$$

如何检验(5.17)(即如何检验(5.9)或(5.10))? 下面的两个定理分别给出了检验统计量及其分布。

定理 5.1 对于单向多元方差分析模型(5.16)，检验假设(5.17)的极大似然比统计量是

$$\Lambda = |E| / |E + B| \tag{5.20}$$

其中 B 和 E 为(5.15)式所示。

证明 设 $\Omega = \{(\mu_1 \cdots \mu_m, \Sigma):\ \Sigma > 0,\ \mu_i$ 任意$\}$. $\Omega_0 = \{(\mu_1 \cdots \mu_m, \Sigma):\ \Sigma > 0, \mu_1 = \cdots = \mu_m\}$. 那么在 Ω 下的似然函数为

$$\begin{aligned} L(\mu_1 \cdots \mu_m, \Sigma) &= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left\{\mathrm{tr}\left[-\frac{1}{2} \times \Sigma^{-1}(X - M)(X - M)'\right]\right\} \\ &= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \prod_{i=1}^{m} \exp\left\{\mathrm{tr}\left[-\frac{1}{2} \times \Sigma^{-1}(X_i - M_i)(X_i - M_i)'\right]\right\} \\ &= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \prod_{i=1}^{m} \exp\left\{\mathrm{tr}\left[-\frac{1}{2} \times \Sigma^{-1}(E_i + n_i(\bar{x}_{i.} - \mu_i)(\bar{x}_i - \mu_i))\right]\right\} \end{aligned}$$

从而当 $\boldsymbol{\mu}_i$ 和 $\boldsymbol{\Sigma}$ 取

$$\hat{\boldsymbol{\mu}}_i = \bar{\boldsymbol{x}}_i, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{m} E_i = \frac{1}{n} E$$

时，上述似然函数达到最大，即

$$\max_{\Omega} L(\boldsymbol{\mu}_1 \cdots \boldsymbol{\mu}_m, \Sigma) = L(\hat{\boldsymbol{\mu}}_1 \cdots \hat{\boldsymbol{\mu}}_m, \hat{\Sigma})$$
$$= (2\pi)^{-\frac{np}{2}} \left| \frac{1}{n} E \right|^{-\frac{n}{2}} \exp\left(-\frac{np}{2}\right)$$

在 Ω_0 下(即 H_0 成立时)其似然函数为

$$L(\boldsymbol{\mu}_1 \cdots \boldsymbol{\mu}_m, \Sigma) = L(\boldsymbol{\mu} \cdots \boldsymbol{\mu}, \Sigma)$$
$$= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left\{\operatorname{tr}\left[-\frac{1}{2}\Sigma^{-1}(X - M_0)(X - M_0)'\right]\right\}$$
$$= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left\{\operatorname{tr}\left[-\frac{1}{2}\right.\right.$$
$$\left.\left. \times \Sigma^{-1}(Q + n(\bar{\boldsymbol{x}}_{..} - \boldsymbol{\mu})(\bar{\boldsymbol{x}}_{..} - \boldsymbol{\mu})')\right]\right\}$$

从而当 $\boldsymbol{\mu}$ 和 Σ 取

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}_{..}, \quad \hat{\Sigma} = \frac{1}{n} Q$$

时，上面的似然函数达到最大，即

$$\max_{\Omega_0} L(\boldsymbol{\mu}_1 \cdots \boldsymbol{\mu}_m, \Sigma) = (2\pi)^{-\frac{np}{2}} \left| \frac{1}{n} Q \right|^{-\frac{n}{2}} \exp\left(\frac{np}{2}\right)$$

故似然比统计量为

$$\Lambda^* = \frac{\max\limits_{\Omega_0} L(\boldsymbol{\mu}_1 \cdots \boldsymbol{\mu}_m, \Sigma)}{\max\limits_{\Omega} L(\boldsymbol{\mu}_1 \cdots \boldsymbol{\mu}_m, \Sigma)} = \left(\frac{|E|}{|Q|}\right)^{\frac{n}{2}}$$

它是等价于

$$\Lambda = |E|/|Q| = |E|/|E + B| \qquad \blacksquare$$

定理 5.2 在(5.17)中的 H_0 成立时，则有

$$\Lambda = |E|/|E + B| \sim \Lambda(p, n - m, m - 1) \tag{5.21}$$

证明　在 H_0 成立时，$X \sim N_{p,n}(M_0, \Sigma, I_n)$。由于 C 是投影阵，且 $\text{rk}C = n-1$，$M_0C = 0$，根据第三章推论 2.5，就有

$$Q = XCX' \sim W_p(n-1, \Sigma)$$

同样，由 $E = \sum_{i=1}^{m} E_i = \sum_{i=1}^{m} X_iC_iX_i' = X\tilde{C}X'$，$\tilde{C} = \text{diag}(C_1, \cdots, C_m)$ 也是投影阵，且 $\text{rk}\tilde{C} = \sum_{i=1}^{m}(n_i - 1) = n - m$，$M\tilde{C} = 0$。于是有

$$E = X\tilde{C}X' \sim W_p(n-m, \Sigma)$$

由(5.15)式知

$$B = Q - E = XCX' - X\tilde{C}X' = X(C - \tilde{C})X'$$

由 $C\tilde{C} = \tilde{C}C = \tilde{C}$ 可推出 $C - \tilde{C}$ 也是投影阵，且 $M_0(C-\tilde{C}) = 0$，$\text{rk}(C-\tilde{C}) = \text{tr}(C-\tilde{C}) = (n-1)-(n-m) = m-1$，$(C-\tilde{C})\cdot\tilde{C} = 0$，从而由第三章定理 2.9 知道 E 和 B 相互独立，按 Wilks-Λ 统计量的定义便证明了定理。▌

由上可知检验(5.17)(或(5.9),(5.10))的步骤如下:

(1) 分别计算 B 和 $E(Q = B + E)$。

(2) 求出 $\Lambda = |E|/|B+E|$。

(3) 对于给定的显著水平 α，求出 Λ_α 使得 $P(\Lambda(p, n-m, m-1) > \Lambda_\alpha) = \alpha$。如果当 p 和 $m-1$ 为 1 或 2 时，可利用第三章的表 3.1 化为 F 检验，然后查 F 表，否则利用第三章(3.6)式查 χ^2 表近似地求出。

(4) 根据临界值来推断 H_0 是否成立。

最后我们可归成一个方差分析表(表 5.2)。

例 5.1　考查医院所有制对医疗费用的影响，所有制分三种：(1) 私立，(2) 非盈利组织经营，(3) 政府开办。医疗费用由四个分量组成：x_1 = 护理费用，x_2 = 膳食服务费用，x_3 = 设备使用及维修费用，x_4 = 房产管理及洗衣费用。这四个费用都是以每个病人每天开支为基础计算的。三种类型的医院分别观察到的样

表 5.2　单向多元方差分析表

方差来源	离　差　矩　阵　和	自由度	统计量
因素的影响	$B=\sum_{i=1}^{m} n_i(\bar{x}_{i.}-\bar{x}_{..})(\bar{x}_{i.}-\bar{x}_{..})'$	$m-1$	$\Lambda=\lvert E\rvert/\lvert B+E\rvert$
误差	$E=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})(x_{ij}-\bar{x}_{i.})'$	$n-m$	
总和	$Q=B+E=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{..})(x_{ij}-\bar{x}_{..})'$	$n-1$	

本均值向量及样本的协方差矩阵分别如下：

$$\text{私立：}\ n_1=271,\quad \bar{x}_{1.}=\begin{pmatrix}2.066\\0.430\\0.082\\0.360\end{pmatrix}$$

$$S_1=\begin{pmatrix}0.291 & & & \\ -0.001 & 0.011 & & \\ 0.002 & 0.000 & 0.001 & \\ 0.010 & 0.003 & 0.000 & 0.010\end{pmatrix}$$

$$\text{非盈利组织：}\ n_2=138,\quad \bar{x}_{2.}=\begin{pmatrix}2.167\\0.596\\0.124\\0.418\end{pmatrix}$$

$$S_2=\begin{pmatrix}0.561 & & & \\ 0.011 & 0.025 & & \\ 0.001 & 0.004 & 0.005 & \\ 0.037 & 0.007 & 0.002 & 0.019\end{pmatrix}$$

政府开办：$n_3=107$，$\bar{x}_{3\cdot}=\begin{pmatrix}2.273\\0.521\\0.125\\0.383\end{pmatrix}$

$$S_3=\begin{pmatrix}0.261 & & & \\ 0.030 & 0.017 & & \\ 0.003 & 0.000 & 0.004 & \\ 0.013 & 0.006 & 0.013 & 0.015\end{pmatrix}$$

于是可以计算得

$$\bar{x}_{\cdot\cdot}=\frac{1}{n}\sum_{i=1}^{3} n_i\bar{x}_{i\cdot}=\begin{pmatrix}2.136\\0.519\\0.120\\0.380\end{pmatrix}$$

注意到 S_i 是样本协方差矩阵，即

$$S_i=\frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})(x_{ij}-\bar{x}_{i\cdot})'=\frac{1}{n_i-1}E_i$$

所以

$$E=\sum_{i=1}^{3}(n_i-1)S_i=\begin{pmatrix}182.962 & & & \\ 4.408 & 8.200 & & \\ 1.695 & 0.633 & 1.484 & \\ 9.581 & 2.428 & 0.394 & 0.538\end{pmatrix}$$

$$B=\sum_{i=1}^{3}n_i(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})'=\begin{pmatrix}3.475 & & & \\ 1.111 & 1.225 & & \\ 0.821 & 0.453 & 0.235 & \\ 0.584 & 0.610 & 0.230 & 0.304\end{pmatrix}$$

由此计算出

$$\Lambda=|E|/|E+B|=0.7714$$

本例中 $p=4$，$n=n_1+n_2+n_3=516$，$m=3$，所以由定理 5.2

知道 $\Lambda \sim \Lambda\ (4,513,2)$，这里 $m-1=2$。由第三章表 3.1 知

$$F=\frac{509}{4}\cdot\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\sim F(8,1018)$$

此时化成 F 检验，取 $\alpha=0.01$，根据自由度查表得到临界值 $F_{0.99}(8,1018)=2.52$。而

$$F=\frac{509}{4}\cdot\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}=17.63>2.52$$

所以在 $\alpha=0.01$ 之下拒绝原假设 H_0，即三种类型医院的平均医疗费用有显著差异。

如果我们对四个分量 $x_i(i=1,2,3,4)$ 分别用一元方差分析进行处理，看其结论又如何。由于 $E\sim W_p(n-m,\Sigma)$，$B\sim W_p(m-1,\Sigma)$，且相互独立，所以它们的对角元素 $e_{ii}\sim\chi^2(n-m)\cdot\sigma_{ii}$，$b_{ii}\sim\chi^2(m-1)\cdot\sigma_{ii}$ 且相互独立，由此可得一元方差分析的 F 检验统计量

$$F=\frac{b_{ii}/m-1}{e_{ii}/n-m}\sim F(m-1,n-m)$$

于是利用 E 和 B 的对角元素求得

$$F_1=\frac{3.475/2}{182.926/513}=4.8727\qquad F_2=\frac{1.225/2}{8.200/513}=38.3186$$

$$F_3=\frac{0.235/2}{1.484/513}=40.6183\qquad F_4=\frac{0.304/2}{0.538/513}=144.9368$$

查 F 表得临界值 $F_{0.99}(2,513)=6.63$。将 F_i 与 6.63 比较，发现指标 x_2,x_3 与 x_4 有显著差异，而指标 x_1，即护理费用可以认为无显著差异。

从此例看出：当我们使用 Wilks 的 Λ 统计量进行多元方差分析时经常与一元方差分析相结合，多元的具有全面考察的特点，而一元则可能发现各指标间的差异，两者相结合，结论更可靠。

*(二) 因素效应的同时置信区间

当假设 H_0: $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_m$ 被拒绝后，我们感兴趣的是那些均值不等的总体之间的均值的分量差的同时置信区间或者它们的线性组合的同时置信区间。其直观意义是这样：若第 s 个总体的均值向量 $\boldsymbol{\mu}_s$ 不等于第 t 个总体的均值向量 $\boldsymbol{\mu}_t$，我们希望估计出 $\boldsymbol{\mu}_s$ 与 $\boldsymbol{\mu}_t$ 分量之间偏差的情况。在本段中我们介绍 Roy 的方法。为此，先作一些准备工作。

设 $T = (\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_m)$ 是 $p \times m$ 阶矩阵，$\boldsymbol{a}$ 是任一 p 维向量，$\boldsymbol{c}$ 是满足 $\boldsymbol{c}'\mathbf{1}_m = 0$ 的 m 维向量，即 $\boldsymbol{c} = (c_1, \cdots, c_m)'$ 的各分量满足 $\sum_{i=1}^{m} c_i = 0$。如取 $\boldsymbol{a} = (1, 0, \cdots 0)'$，$\boldsymbol{c} = (1, -1, 0, \cdots, 0)'$，则 $\boldsymbol{a}'T\boldsymbol{c} = \mu_{11} - \mu_{21} = \tau_{11} - \tau_{21}$ 正是第一总体和第二总体的均值差的第一个分量。

设

$$\tilde{\boldsymbol{x}}_{ij} = \boldsymbol{x}_{ij} - \boldsymbol{\mu}_i \sim N_p(0, \Sigma) \left(i = 1, 2, \cdots, m, \right.$$

$$\left. j = 1, 2, \cdots, n_i, \quad \sum_{i=1}^{m} n_i = n\right)$$

记 $\tilde{X} = (\tilde{X}_1, \cdots, \tilde{X}_m)$，$\tilde{X}_i = (\tilde{\boldsymbol{x}}_{i1}, \cdots, \tilde{\boldsymbol{x}}_{in_i})$，则

$$\tilde{X} \sim N_{p,n}(0, \Sigma, I_n) \tag{5.22}$$

类似于定理 5.2 的证明(取 $M_0 = 0$，以 $\tilde{X}$ 代替 X) 立刻可得到

$$\begin{aligned}\tilde{E} &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} (\tilde{\boldsymbol{x}}_{ij} - \tilde{\boldsymbol{x}}_{i\cdot})(\tilde{\boldsymbol{x}}_{ij} - \tilde{\boldsymbol{x}}_{i\cdot})' \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \boldsymbol{x}_{i\cdot})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i\cdot})' \\ &= E \sim W_p(n - m, \Sigma)\end{aligned} \tag{5.23}$$

$$\begin{aligned}\tilde{B} &= \sum_{i=1}^{m} n_i(\tilde{\boldsymbol{x}}_{i\cdot} - \tilde{\boldsymbol{x}}_{\cdot\cdot})(\tilde{\boldsymbol{x}}_{i\cdot} - \tilde{\boldsymbol{x}}_{\cdot\cdot})' \\ &= \sum_{i=1}^{m} n_i(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot} - \boldsymbol{\tau}_i)(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot} - \boldsymbol{\tau}_i)'\end{aligned}$$

$$\sim W_p(m-1,\Sigma) \tag{5.24}$$

且 $\tilde{E}$ 和 $\tilde{B}$ 相互独立。

设 η_{max} 是 $|\tilde{B}-\eta\tilde{E}|=0$ 的最大特征根，θ_{max} 是 $|\tilde{B}-\theta(\tilde{B}+\tilde{E})|=0$ 的最大特征根。则由于

$$|\tilde{B}-\theta(\tilde{B}+\tilde{E})| = |(1-\theta)\tilde{B}-\theta\tilde{E}|$$
$$= (1-\theta)^p \left|\tilde{B}-\frac{\theta}{1-\theta}\tilde{E}\right| = 0$$

所以有

$$\eta_{max} = \theta_{max}/1-\theta_{max} \tag{5.25}$$

用 η_α 和 θ_α 分别记 η_{max} 和 θ_{max} 的 α 分位点，若 $P(\eta_{max}<\eta_\alpha)=\alpha$，由于 η_{max} 是 θ_{max} 的增函数（由 $\eta_{max}>0$ 及(5.25)知道 $0<\theta_{max}<1$），所以若 $P(\theta_{max}<\theta_\alpha)=\alpha$，则有 $P\left(\eta_{max}<\dfrac{\theta_\alpha}{1-\theta_\alpha}\right)=\alpha$，因此 $\eta_\alpha=\theta_\alpha/1-\theta_\alpha$。即得到 η_{max} 与 θ_{max} 的 α 分位点之间的关系。

关于 θ_{max} 的分布经前人的研究已制成了 α 分位点表(书末附表 $A5$)。在此对附表 $A5$ 中 $\theta_{max}(p_0, n, m_0, \alpha)$ 的各个参数的意义作一点说明：若 $A\sim W_{p_0}(m_0,\Sigma)$，$B\sim W_{p_0}(n,\Sigma)$，则 $\theta_{max}(p_0, n, m_0, \alpha)$ 正是 $|A-\theta(A+B)|=0$ 的最大特征根 θ_{max} 分布的 α 分位点。由此可知 $|\tilde{B}-\theta(\tilde{B}+\tilde{E})|=0$ 的最大特征根的 α 分位点正是 $\theta_{max}(p,n-m,m-1,\alpha)$。

有了上述的准备工作，下面我们给出 $a'Tc=\sum_{i=1}^{m}c_i a'\mu_i$ 的同时置信区间。

定理 5.3 对于单向多元方差分析模型(5.11)，当 $n-m>p$ 时，对于任一 p 维向量 a 和满足 $c'1_m=0$ 的 m 维向量 c，$a'Tc=\sum_{i=1}^{m}c_i a'\mu_i$ 的 $1-\alpha$ 同时置信区间为

$$\left(\sum_{i=1}^{m} c_i \boldsymbol{a}'\bar{\boldsymbol{x}}_{i\cdot} \pm \left[\frac{\theta_\alpha}{1-\theta_\alpha} \cdot \boldsymbol{a}'\tilde{E}\boldsymbol{a} \cdot \left(\sum_{i=1}^{m} \frac{c_i^2}{n_i}\right)\right]^{\frac{1}{2}}\right) \tag{5.26}$$

其中 $\theta_\alpha = \theta_{\max}(p, n-m, n-1, \alpha)$，即 $\theta_{\max}$ 的 α 分位点。

证明　与 $T = (\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_m)$ 相对应，我们记 $\hat{T} = (\bar{\boldsymbol{x}}_{1\cdot}, \cdots, \bar{\boldsymbol{x}}_{m\cdot})$，$V = (\boldsymbol{\tau}_1, \cdots, \boldsymbol{\tau}_m)$。则有

$$E\hat{T} = T,\ E(\boldsymbol{a}'\hat{T}\boldsymbol{c}) = \boldsymbol{a}'T\boldsymbol{c},\ T - \boldsymbol{\mu}\mathbf{1}_m' = V \tag{5.27}$$

利用第一章引理 3.1 中的 Cauchy-Schwarz 不等式，即对任意二个 p 维向量 $\boldsymbol{b}$ 及 $\boldsymbol{d}$ 有

$$(\boldsymbol{b}'\boldsymbol{d})^2 \leqslant (\boldsymbol{b}'\boldsymbol{b}) \cdot (\boldsymbol{d}'\boldsymbol{d}) \tag{5.28}$$

令

$$\boldsymbol{d} = (n_1^{-\frac{1}{2}}c_1, \cdots, n_m^{-\frac{1}{2}}c_m)' = [\mathrm{diag}(n_1, \cdots, n_m)]^{-\frac{1}{2}} \cdot \boldsymbol{c} \tag{5.29}$$

$$\begin{aligned}\boldsymbol{b} &= (n_1^{\frac{1}{2}}\boldsymbol{a}'(\bar{\boldsymbol{x}}_{1\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot} - \boldsymbol{\tau}_1), \cdots, n_m^{\frac{1}{2}}\boldsymbol{a}'(\bar{\boldsymbol{x}}_{m\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot} - \boldsymbol{\tau}_m))' \\ &= [\mathrm{diag}(n, \cdots, n_m)]^{\frac{1}{2}}(\boldsymbol{a}'(\hat{T} - V - \bar{\boldsymbol{x}}_{\cdot\cdot}\mathbf{1}_m'))'\end{aligned} \tag{5.30}$$

注意到(5.27)，(5.24)式及 $\mathbf{1}_m'\boldsymbol{c} = 0$，从而有

$$\begin{aligned}\boldsymbol{b}'\boldsymbol{d} &= \boldsymbol{a}'(\hat{T} - V - \bar{\boldsymbol{x}}_{\cdot\cdot}\mathbf{1}_m')\boldsymbol{c} = \boldsymbol{a}'[(\hat{T} - V)\boldsymbol{c} - \bar{\boldsymbol{x}}_{\cdot\cdot}\mathbf{1}_m'\boldsymbol{c}] \\ &= \boldsymbol{a}'(\hat{T} - T)\boldsymbol{c} = \sum_{i=1}^{m} c_i\boldsymbol{a}'(\bar{\boldsymbol{x}}_{i\cdot} - \boldsymbol{\mu}_i)\end{aligned} \tag{5.31}$$

$$\boldsymbol{d}'\boldsymbol{d} = \boldsymbol{c}'\mathrm{diag}(n_1, \cdots, n_m)^{-1}\boldsymbol{c} = \sum_{i=1}^{m} \frac{c_i^2}{n_i} \tag{5.32}$$

$$\begin{aligned}\boldsymbol{b}'\boldsymbol{b} &= \sum_{i=1}^{m} n_i[\boldsymbol{a}'(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot} - \boldsymbol{\tau}_i)]^2 \\ &= \boldsymbol{a}'\left[\sum_{i=1}^{m} n_i(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot} - \boldsymbol{\tau}_i)(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}}_{\cdot\cdot} - \boldsymbol{\tau}_i)'\right]\boldsymbol{a} \\ &= \boldsymbol{a}'\tilde{B}\boldsymbol{a}\end{aligned} \tag{5.33}$$

将(5.31)～(5.33)代入(5.28)得到

$$[\boldsymbol{a}'(\hat{T} - T)\boldsymbol{c}]^2 = \left[\sum_{i=1}^{m} c_i\boldsymbol{a}'(\bar{\boldsymbol{x}}_{i\cdot} - \boldsymbol{\mu}_i)\right]^2$$

$$\leqslant a'\widetilde{B}a \cdot \sum_{i=1}^{m} \frac{c_i^2}{n_i} \tag{5.34}$$

由第一章推论 3.2 知道，对于任一 p 维向量 a 有

$$\max_{a} \frac{a'\widetilde{B}a}{a'\widetilde{E}a} \leqslant \eta_{\max}$$

其中 $\eta_{\max}$ 是 $\widetilde{B}$ 相对于 $\widetilde{E}$ 的最大特征根，即 $|\widetilde{B}-\eta\widetilde{E}|=0$ 的最大特征根。在(5.34)式两边同除 $a'\widetilde{E}a$，

$$\frac{[a'(\hat{T}-T)c]^2}{a'\widetilde{E}a} \leqslant \sum_{i=1}^{m} \frac{c_i^2}{n_i}\left(\frac{a'\widetilde{B}a}{a'\widetilde{E}a}\right) \leqslant \sum_{i=1}^{m} \frac{c_i^2}{n_i} \cdot \eta_{\max}$$

从而有

$$[a'(\hat{T}-T)c]^2 = \left[\sum_{i=1}^{m} c_i a'(\bar{x}_{i\cdot}-\mu_i)\right]^2$$

$$\leqslant \left(\sum_{i=1}^{m} \frac{c_i^2}{n_i}\right) \cdot a'\widetilde{E}a \cdot \eta_{\max}$$

或者写成

$$\left|\sum_{i=1}^{m} c_i a'\bar{x}_{i\cdot} - \sum_{i=1}^{m} c_i a'\mu_i\right| \leqslant \left[\left(\sum_{i=1}^{m} \frac{c_i^2}{n_i}\right) \cdot a'\widetilde{E}a \cdot \eta_{\max}\right]^{\frac{1}{2}} \tag{5.35}$$

对于给定的显著水平 α，可得 η_α 使得

$$P\left\{\left(\sum_{i=1}^{m} \frac{c_i^2}{n_i} \cdot a'\widetilde{E}a \cdot \eta_{\max}\right)^{\frac{1}{2}} \leqslant \left(\sum_{i=1}^{m} \frac{c_i^2}{n_i} \cdot a'\widetilde{E}a \cdot \eta_\alpha\right)^{\frac{1}{2}}\right\} = 1-\alpha$$

由(5.35)式知道事件

$$\left\{\left(\sum_{i=1}^{m} \frac{c_i^2}{n_i} \cdot a'\widetilde{E}a \cdot \eta_{\max}\right) \leqslant \left(\sum_{i=1}^{m} \frac{c_i^2}{n_i} \cdot a'\widetilde{E}a \cdot \eta_\alpha\right)^{\frac{1}{2}}\right\}$$

发生必导致事件

$$\left\{\left|\sum_{i=1}^{m} c_i a'\bar{x}_{i\cdot} - \sum_{i=1}^{m} c_i a'\mu_i\right| \leqslant \left(\sum_{i=1}^{m} \frac{c_i^2}{n_i} \cdot a'\widetilde{E}a \cdot \eta_\alpha\right)^{\frac{1}{2}}\right\}$$

发生，其概率不会小于 $1-\alpha$。注意到 $\eta_\alpha=\dfrac{\theta_\alpha}{1-\theta_\alpha}$，这样我们就得到 $\sum_{i=1}^{m} c_i a'\mu_i$ 的 $1-\alpha$ 同时置信区间为(5.26)式。▮

例 5.2 在例 5.1 中我们看到医院的平均费用与其类型有关，应用定理 5.3 我们可以比较费用差异的大小。譬如比较私立医院与政府开设的医院的设备使用和维修费 x_3，即要求出 $\mu_{13}-\mu_{33}$ 的置信区间。为此取 $c=(1,0,-1)'$，$a=(0,0,1,0)'$ 于是有

$$\sum_{i=1}^{3} c_i a'\bar{x}_{i\cdot}=a'(\bar{x}_{1\cdot}-\bar{x}_{3\cdot})=-0.043,$$

$$\sum_{i=1}^{3}\frac{c_i^2}{n_i}=0.01304,\quad a'\tilde{E}a=1.484$$

由 $p=4$，$n=513$，$m=2$ 查得 $\theta_{0.05}=\theta_{\max}(4,\ 513,\ 2,\ 0.95)=0.025$(注意书末附表 $A5$ 对于 n 只列到 101，当 $n>101$ 时要查更详尽的表)。因此 $\mu_{13}-\mu_{33}$ 的 0.95 置信区间为

$$\left(-0.043\pm\left[\frac{0.025}{1-0.025}\times 1.484\times 0.01304\right]^{\frac{1}{2}}\right)$$

$$=(-0.065,-0.021)$$

这说明，政府医院比私立医院在设备使用和维修费方面要高出 0.021 到 0.065。

用同样方法可求得 $\mu_{23}-\mu_{33}$ 的 0.95 置信区间为 $(-0.026,0.024)$，而 0 属于该区间内，这说明非盈利组织办的医院与政府办的医院的设备使用与维修费用无显著差异，这正是对例 5.1 的讨论的更细微的补充。

(三) 双向分类

现在考虑两个因素 A 与 B 以及它们的交互作用 $A\times B$ 是否对试验结果有显著影响的问题。设因素 A 取 m 个水平：$A_i(i=1,2,\cdots,m)$，因素 B 取 s 个水平：$B_j(j=1,2,\cdots,s)$，共有 $m\times s$ 个配合，由于要考虑交互作用，因而必须在每一个配合下做重

复试验，并设重复 t 次（$t\geqslant 2$）。每次试验结果是一个 p 维向量 $\boldsymbol{x}$，用 $\boldsymbol{x}_{ijk}$ 表示因素 A 在第 i 水平，因素 B 在第 j 水平下所做的第 k 次试验结果，考虑下列的模型：

$$\begin{cases}\boldsymbol{x}_{ijk}=\boldsymbol{\mu}_{ij}+\boldsymbol{\varepsilon}_{ijk}\\ \boldsymbol{\varepsilon}_{ijk}\sim N_p(0,\Sigma)\text{ 且相互独立 }(i=1,2,\cdots,m,\\ \qquad\qquad j=1,2,\cdots,s,\ k=1,\cdots,t)\end{cases}\tag{5.36}$$

为了今后便于讨论和解释，我们令

$$\boldsymbol{\mu}=\frac{1}{ms}\sum_{i=1}^{m}\sum_{j=1}^{s}\boldsymbol{\mu}_{ij},\ \boldsymbol{\mu}_{i\cdot}=\frac{1}{s}\sum_{j=1}^{s}\boldsymbol{\mu}_{ij},\ \boldsymbol{\mu}_{\cdot j}=\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{\mu}_{ij}$$

$$\boldsymbol{\tau}_i=\boldsymbol{\mu}_{i\cdot}-\boldsymbol{\mu},\ \boldsymbol{\beta}_j=\boldsymbol{\mu}_{\cdot j}-\boldsymbol{\mu},$$

$$\boldsymbol{\gamma}_{ij}=\boldsymbol{\mu}_{ij}-(\boldsymbol{\mu}+\boldsymbol{\tau}_i+\boldsymbol{\beta}_j)=\boldsymbol{\mu}_{ij}-\boldsymbol{\mu}_{i\cdot}-\boldsymbol{\mu}_{\cdot j}+\boldsymbol{\mu}$$

其中 $\boldsymbol{\tau}_i,\boldsymbol{\beta}_j$ 分别表示因素 A 的第 i 水平的效应和因素 B 的第 j 水平的效应，而 $\boldsymbol{\gamma}_{ij}$ 表示它们的交互作用的效应。如果不存在交互作用，则 $\boldsymbol{\gamma}_{ij}=0$。由上所述不难验证有

$$\sum_{i=1}^{m}\boldsymbol{\tau}_i=0,\ \sum_{j=1}^{s}\boldsymbol{\beta}_j=0,\ \sum_{i=1}^{m}\boldsymbol{\gamma}_{ij}=0,\ \sum_{j=1}^{m}\boldsymbol{\gamma}_{ij}=0\tag{5.37}$$

成立。此时我们可将模型(5.36)改写为

$$\begin{cases}\boldsymbol{x}_{ijk}=\boldsymbol{\mu}+\boldsymbol{\tau}_i+\boldsymbol{\beta}_j+\boldsymbol{\gamma}_{ij}+\boldsymbol{\varepsilon}_{ijk}\\ \boldsymbol{\varepsilon}_{ijk}\sim N_p(0,\Sigma)\text{ 且相互独立 }(i=1,2,\cdots,m,\\ \qquad\qquad j=1,2,\cdots,s,\ k=1,2,\cdots,t)\end{cases}\tag{5.38}$$

由模型(5.38)可知，考查因素 A，因素 B 及它们的交互作用对试验结果是否有显著影响就等价于检验如下三个假设：

$$H_{01}:\boldsymbol{\tau}_1=\boldsymbol{\tau}_2=\cdots=\boldsymbol{\tau}_m=0\quad H_{11}:\ \boldsymbol{\tau}_i\text{ 不全为 }0\tag{5.39}$$

$$H_{02}:\boldsymbol{\beta}_1=\boldsymbol{\beta}_2=\cdots=\boldsymbol{\beta}_s=0\quad H_{12}:\ \boldsymbol{\beta}_j\text{ 不全为 }0\tag{5.40}$$

$$H_{03}:\ \boldsymbol{\gamma}_{ij}=0\ (\text{一切 } i \text{ 和 } j)\quad H_{13}:\ \boldsymbol{\gamma}_{ij}\text{ 不全为 }0\tag{5.41}$$

下面直观地介绍检验(5.39)～(5.41)的问题，而不作具体的证明。对于每个观察结果作下列的分解：

$$\boldsymbol{x}_{ijk}=\bar{\boldsymbol{x}}+(\bar{\boldsymbol{x}}_{i\cdot}-\bar{\boldsymbol{x}})+(\bar{\boldsymbol{x}}_{\cdot j}-\bar{\boldsymbol{x}})$$
$$+(\bar{\boldsymbol{x}}_{ij\cdot}-\bar{\boldsymbol{x}}_{i\cdot}-\bar{\boldsymbol{x}}_{\cdot j}+\bar{\boldsymbol{x}})+(\boldsymbol{x}_{ijk}-\bar{\boldsymbol{x}}_{ij\cdot})\tag{5.42}$$

其中

$$\bar{x}=\frac{1}{mst}\sum_{i=1}^{m}\sum_{j=1}^{s}\sum_{k=1}^{t}x_{ijk} \quad \text{为总平均向量}$$

$$\bar{x}_{i\cdot}=\frac{1}{st}\sum_{j=1}^{s}\sum_{k=1}^{t}x_{ijk} \quad \text{为在 } A_i \text{ 下的平均向量}$$

$$\bar{x}_{\cdot j}=\frac{1}{mt}\sum_{i=1}^{m}\sum_{k=1}^{t}x_{ijk} \quad \text{为在 } B_j \text{ 下的平均向量}$$

$$\bar{x}_{ij\cdot}=\frac{1}{t}\sum_{k=1}^{t}x_{ijk} \quad \text{为在 } (A_iB_j) \text{ 配合下的平均向量}$$

由于 $E\bar{x}=\mu$，所以 $\bar{x}$ 反映了平均效应；$E(\bar{x}_{i\cdot}-\bar{x})=\alpha_i$ 所以 $\bar{x}_{i\cdot}-\bar{x}$ 反映了 A_i 的效应；$E(\bar{x}_{\cdot j}-\bar{x})=\beta_j$，所以 $\bar{x}_{\cdot j}-\bar{x}$ 反映了 B_j 的效应；$E(\bar{x}_{ij\cdot}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j}+\bar{x})=\gamma_{ij}$，所以 $\bar{x}_{ij\cdot}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j}+\bar{x}$ 反映了 A_i 与 B_j 的交互作用效应；而 $E(x_{ijk}-\bar{x}_{ij\cdot})=0$，所以 $x_{ijk}-\bar{x}_{ij\cdot}$ 反映了误差。因此我们令

$$S_A=\sum_{i=1}^{m}st(\bar{x}_{i\cdot}-\bar{x})(\bar{x}_{i\cdot}-\bar{x})'$$

$$S_B=\sum_{j=1}^{s}mt(\bar{x}_{\cdot j}-\bar{x})(\bar{x}_{\cdot j}-\bar{x})'$$

$$S_{A\times B}=\sum_{i=1}^{m}\sum_{j=1}^{s}t(\bar{x}_{ij\cdot}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j}+\bar{x})\times(\bar{x}_{ij\cdot}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j}+\bar{x})'$$

$$S_e=\sum_{i=1}^{m}\sum_{j=1}^{s}t(x_{ijk}-\bar{x}_{ij\cdot})(x_{ijk}-\bar{x}_{ij\cdot})'$$

$$Q=\sum_{i=1}^{m}\sum_{j=1}^{s}\sum_{k=1}^{t}(x_{ijk}-\bar{x})(x_{ijk}-\bar{x})'$$

仿照单向分类一样，称 S_A 为因素 A 的效应"平方和"矩阵，S_B 为因素 B 的效应"平方和"矩阵，$S_{A\times B}$ 为交互作用效应"平方和"矩阵，S_e 为误差"平方和"矩阵，Q 为总离差"平方和"矩阵。

通过(5.42)的分解式容易验证

$$Q = S_A + S_B + S_{A\times B} + S_e \tag{5.43}$$

而 $S_A, S_B, S_{A\times B}, S_e$ 和 Q 都是 x_{ijk} 的二次型。在 S_A 中有 m 个变数，存在一个线性关系式 $\sum_{i=1}^{m}(\bar{x}_{i\cdot} - \bar{x}) = 0$，故它有自由度 $m-1$，类似地 S_B 有自由度 $s-1$，$S_{A\times B}$ 有自由度 $(m-1)(s-1)$，S_e 有自由度 $ms(t-1)$，Q 有自由度 $mst-1$。

检验(5.39)～(5.41)的检验统计量分别为

$$\Lambda_A = |S_e|/|S_e + S_A| \sim \Lambda(p, ms(t-1), m-1) \tag{5.44}$$

$$\Lambda_B = |S_e|/|S_e + S_B| \sim \Lambda(p, ms(t-1), s-1) \tag{5.45}$$

$$\Lambda_{A\times B} = |S_e|/|S_e + S_{A\times B}| \sim \Lambda(p, ms(t-1), (m-1)(s-1)) \tag{5.46}$$

以上可归纳成一个方差分析表(表 5.3)。

表 5.3　双向多元方差分析表

方差来源	离差和矩阵	自 由 度	统计量
因 素 A	$S_A = \sum_i st(\bar{x}_{i\cdot\cdot} - \bar{x})(\bar{x}_{i\cdot\cdot} - \bar{x})'$	$m-1$	$\|S_e\|/\|S_e + S_A\|$
因 素 B	$S_B = \sum_j mt(\bar{x}_{\cdot j\cdot} - \bar{x})(\bar{x}_{\cdot j\cdot} - \bar{x})'$	$s-1$	$\|S_e\|/\|S_e + S_B\|$
交互作用	$S_{A\times B} = \sum_i \sum_j t(\bar{x}_{ij\cdot} - \bar{x}_{i\cdot\cdot} - \bar{x}_{\cdot j\cdot} + \bar{x})(\bar{x}_{ij\cdot} - \bar{x}_{i\cdot\cdot} - \bar{x}_{\cdot j\cdot} + \bar{x})'$	$(m-1)(s-1)$	$\|S_e\|/\|S_e + S_{A\times B}\|$
误 差	$S_e = \sum_i \sum_j t(\bar{x}_{ijk} - \bar{x}_{ij\cdot})(\bar{x}_{ijk} - \bar{x}_{ij\cdot})'$	$ms(t-1)$	
总 和	$Q = \sum_i \sum_j \sum_k (x_{ijk} - \bar{x})(x_{ijk} - \bar{x})'$	$mst-1$	

若由于实验条件限制，在 A 和 B 的各水平配合下只能做一次试验，即总共做 ms 次试验，那么就无法了解它们的交互作用，此时只能考虑模型

$$\begin{cases} x_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \\ \epsilon_{ij} \sim N_p(0, \Sigma) \text{ 且相互独立 } (i = 1, 2, \cdots, m. \\ \qquad\qquad\qquad\qquad\qquad j = 1, 2, \cdots, s) \end{cases} \tag{5.47}$$

类似于(5.42)的分解式为

$$x_{ij} = \bar{x} + (\bar{x}_{i\cdot} - \bar{x}) + (\bar{x}_{\cdot j} - \bar{x}) + (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}) \tag{5.48}$$

其中

$$\bar{x} = \frac{1}{ms} \sum_i \sum_j x_{ij}, \quad \bar{x}_{i\cdot} = \frac{1}{s} \sum_j x_{ij}, \quad x_{\cdot j} = \frac{1}{m} \sum_i x_{ij}$$

令

$$S_A = \sum_i s(\bar{x}_{i\cdot} - \bar{x})(\bar{x}_{i\cdot} - \bar{x})' \quad (\text{自由度为 } m - 1) \tag{5.49}$$

$$S_B = \sum_j m(\bar{x}_{\cdot j} - \bar{x})(\bar{x}_{\cdot j} - \bar{x})' \quad (\text{自由度为 } s - 1) \tag{5.50}$$

$$S_e = \sum_i \sum_j (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})(x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})' \quad (\text{自由度为 } (m - 1)(s - 1)) \tag{5.51}$$

$$Q = \sum_i \sum_j (x_{ij} - \bar{x})(x_{ij} - \bar{x})' \quad (\text{自由度为 } ms - 1)$$

通过分解式(5.48)，容易验证

$$Q = S_A + S_B + S_e \tag{5.52}$$

此时要检验的假设是(5.39)和(5.40)。检验的统计量为

$$\Lambda_A = |S_e| / |S_e + S_A| \sim \Lambda(p, (m - 1)(s - 1), m - 1) \tag{5.53}$$

$$\Lambda_B = |S_e| / |S_e + S_B| \sim \Lambda(p, (m - 1)(s - 1), s - 1) \tag{5.54}$$

其中 S_A, S_B 和 S_e 为(5.49)～(5.51)所确定。类似地，可列成一个双向无交互作用的多元方差分析表。

例 5.3 我们把四个不同的谷种随机分成四个组在 20 块土地上按五种不同的种植方式进行培育（即不同的肥料用量和不同的田间管理措施），在移植后六个星期对二个指标进行测量，一个是

植株的高度 x_1，另一个是植株的分蘖数 x_2。 表 5.4 给出了有关的数据。我们希望了解不同的品种和不同的种植方式对两个指标是否有显著影响。这正是双向的多元方差分析问题。 由表 5.4 的数据知道不同品种和不同的种植方式的各种配合只做了一次试验，故不考虑交互作用，按模型(5.47)进行处理。由表(5.4)中的数据及(5.49)～(5.51)计算得

$$S_A = \begin{pmatrix} 16.5 & 12.4 \\ 12.4 & 5.8 \end{pmatrix} \quad (\text{自由度为 } m-1=3)$$

$$S_B = \begin{pmatrix} 31.7 & -0.5 \\ -0.5 & 1.0 \end{pmatrix} \quad (\text{自由度为 } s-1=4)$$

$$S_e = \begin{pmatrix} 157.4 & -14.9 \\ -14.9 & 18.2 \end{pmatrix} \quad (\text{自由度为 } (m-1)(s-1)=12)$$

$$Q = S_A + S_B + S_e = \begin{pmatrix} 354.2 & -3 \\ -3 & 25 \end{pmatrix} \quad (\text{自由度为 } ms-1=19)$$

表 5.4　水稻品种试验数据

品种 / 数据 / 方式	品种 A		品种 B		品种 C		品种 D	
	x_1	x_2	x_1	x_2	x_1	x_2	x_1	x_2
(一)	58	4	49	7	54	6	58	4
(二)	62	6	55	6	52	6	55	4
(三)	60	7	48	4	51	7	51	5
(四)	54	6	52	4	58	5	49	6
(五)	58	6	49	7	61	6	52	4

检验品种对两个指标是否有显著影响，即要检验

$$H_{01}:\ \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0 \quad H_{11}:\ \tau_i \text{ 不全为 } 0$$

由统计量(5.53)计算得

$$\Lambda_A = |S_e| / |S_e + S_A| = 0.34 \quad (\Lambda_A \sim \Lambda(2,12,3))$$

由第三章表 3.1 知

$$F=\frac{11}{3}\cdot\frac{1-\sqrt{\Lambda_A}}{\sqrt{\Lambda_A}}\sim F(6,22)$$

并由 $\Lambda_A=0.34$ 计算得 $F=2.62$。查得临界值 $F_{0.95}(6,22)=2.55$，显然 $F>2.55$，故拒绝 H_{01}，即认为品种对两项指标有显著影响。

检验种植方式对两项指标是否有显著影响，即要检验

$$H_{02}:\ \beta_1=\beta_2=\cdots=\beta_5=0,\quad H_{12}:\beta_j \text{ 不全为 } 0$$

由统计量(5.54)计算得

$$\Lambda_B=|S_e|/|S_e+S_A|=0.778\quad(\Lambda_B\sim\Lambda(2,12,4))$$

而

$$F=\frac{11}{4}\cdot\frac{1-\sqrt{\Lambda_B}}{\sqrt{\Lambda_B}}\sim F(8,22)$$

并由 $\Lambda_B=0.778$ 计算得 $F=0.367$，临界值为 $F_{0.95}(8,22)=3.45$ 显然 $F<3.45$。故不能拒绝 H_{02}，即各种种植方式无显著差异。

习 题 六

6.1 设有多元正态线性回归模型 $X\sim N_{p,n}(\Theta Z,\Sigma,I_n)$ 试用微分法求出 Θ 的极大似然估计必定满足正规方程

$$XZ'=(ZZ')\Theta.$$

[提示：令 $L(\Theta)$ 是似然函数，由 $dL(\Theta)=0$ 可得出正规方程]

6.2 对于多元线性回归模型 $X=\Theta Z+\varepsilon$，证明 Θ 的 LS 估计 $\hat{\Theta}$ 具有以下性质.

① 使广义方差 $|(X-\Theta Z)(X-\Theta Z)'|$ 达到最小

② 在矩阵非负定意义下，对于任意 Θ 成立

$$(X-\hat{\Theta}Z)(X-\hat{\Theta}Z)'\leqslant(X-\Theta Z)(X-\Theta Z)'$$

③ 使得 $(X-\Theta Z)(X-\Theta Z)'$ 的迹达到最小。

6.3 设 $X\sim N_{p,n}(\Theta Z,\Sigma,I_n)$ $(n>p)$ $(\Sigma>0)$，且 $\mathrm{rk}Z=q+1$，$X_{p\times n}$

随机矩阵. $\Theta_{p\times(q+1)}$ 回归参数矩阵, $Z_{(q+1)\times n}$ 设计矩阵. 设 G 是 $(q+1)\times(q+1)$ 阶非奇矩阵,且 $W=GZ$, 则新的回归参数矩阵为 $\Gamma=\Theta G^{-1}$. 从而 $X\sim N_{p,n}(\Gamma\cdot W,\Sigma,I_n)$. 证明:

① 若 $\hat{\Theta}$ 是 Θ 的极大似然估计,则 Γ 的极大似然估计是 $\hat{\Gamma}=\hat{\Theta}G^{-1}$

② $\hat{\Gamma}WW'\hat{\Gamma}'=\hat{\Theta}(ZZ')\hat{\Theta}'$,从而 $\hat{\Sigma}$(Σ 极大似然估计)不变.

③ $\hat{\Gamma}$ 与 $\hat{\Sigma}$ 相互独立.

④ $\hat{\Gamma}\sim N_{p,q+1}(\Gamma,\Sigma,(G')^{-1}\cdot(ZZ')^{-1}G^{-1})$.

6.4 设多元正态线性模型 $X\sim N_{p,n}(\Theta Z,\Sigma,I_n)$.

其中 $\Theta_{p\times 2q}=(\Theta_1,\Theta_2)$, $Z_{2q\times n}=\begin{pmatrix}Z_1\\ Z_2\end{pmatrix}$ Θ_1 和 Θ_2 都是 $p\times q$ 阶矩阵 Z_1, Z_2 都是 $q\times n$ 阶矩阵. 写出检验假设 $H_0:\Theta_1=\Theta_2$, $H_1:\Theta_1\neq\Theta_2$ 的似然比检验统计量.

6.5 为了建立地面坐标 (x,y) 和卫星扫描图坐标 (u,v) 的变换公式,我们采用多元线性回归模型建立公式(回归方程)

$$\begin{cases}\hat{u}=\hat{\beta}_{01}+\hat{\beta}_{11}x+\hat{\beta}_{12}y\\ \hat{v}=\hat{\beta}_{02}+\hat{\beta}_{12}x+\hat{\beta}_{22}y.\end{cases}$$

现测定了 9 个控制点,数据如下

点序号		1	2	3	4
地理坐标	x	100.38	99.70	92.20	87.60
	y	53.24	51.50	50.50	52.40
扫描坐标	u	226	250	281	272
	v	644	640	517	425

点序号		5	6	7	8	9
地理坐标	x	87.17	88.70	92.75	100.60	90.05
	y	59.00	59.70	65.20	62.40	58.63
扫描坐标	u	194	180	105	115	250
	v	385	401	448	599	462

① 求出回归参数矩阵: $\beta=\begin{pmatrix}\beta_{01} & \beta_{11} & \beta_{12}\\ \beta_{02} & \beta_{12} & \beta_{22}\end{pmatrix}$ 的 LS 估计

② 对 $\alpha = 0.05$. 检验 x, y 对 u, v 有无显著影响?

③ 对于新的坐标 $(x_0, y_0) = (86, 58)'$，预测其相应的扫描图的坐标，求出 (u_0, v_0) 的 0.95 置信椭圆。

6.6 对三种不同的处理方式重复进行试验，分别观察了欲考察的两个指标的值如下：

处理 1: $\begin{pmatrix}6\\7\end{pmatrix}, \begin{pmatrix}5\\9\end{pmatrix}, \begin{pmatrix}8\\6\end{pmatrix}, \begin{pmatrix}4\\9\end{pmatrix}, \begin{pmatrix}7\\9\end{pmatrix}$

处理 2: $\begin{pmatrix}3\\3\end{pmatrix}, \begin{pmatrix}1\\6\end{pmatrix}, \begin{pmatrix}2\\3\end{pmatrix}$

处理 3: $\begin{pmatrix}2\\3\end{pmatrix}, \begin{pmatrix}5\\1\end{pmatrix}, \begin{pmatrix}3\\1\end{pmatrix}, \begin{pmatrix}2\\3\end{pmatrix}$.

① 根据单向分类多元方差分析表进行计算。

② 计算检验统计量 Λ 的值，并分别利用 F 分布以及渐近的 χ^2 分布对检验处理效应是否相等（$\alpha = 0.01$）进行推断，并比较两者结论是否一致。

6.7 要比较 5 个品种大麦产量，用连续二年的单产量 $x = (x_1, x_2)'$ 作为考查指标，在三个不同地区的产量作为三次重复，其资料如下：

连续二年产量 品种 / 重复	A_1	A_2	A_3	A_4	A_5
1	$\begin{pmatrix}81\\81\end{pmatrix}$	$\begin{pmatrix}105\\82\end{pmatrix}$	$\begin{pmatrix}120\\80\end{pmatrix}$	$\begin{pmatrix}110\\87\end{pmatrix}$	$\begin{pmatrix}98\\84\end{pmatrix}$
2	$\begin{pmatrix}147\\100\end{pmatrix}$	$\begin{pmatrix}142\\116\end{pmatrix}$	$\begin{pmatrix}151\\112\end{pmatrix}$	$\begin{pmatrix}192\\148\end{pmatrix}$	$\begin{pmatrix}146\\108\end{pmatrix}$
3	$\begin{pmatrix}120\\99\end{pmatrix}$	$\begin{pmatrix}121\\62\end{pmatrix}$	$\begin{pmatrix}124\\96\end{pmatrix}$	$\begin{pmatrix}141\\126\end{pmatrix}$	$\begin{pmatrix}125\\76\end{pmatrix}$

① 对于 $\alpha = 0.05$，检验各品种之间的产量是否有显著差异。

*② 对于 $\alpha = 0.05$，给出品种 A_1 和品种 A_4 的均值向量差的同时置信区间。

6.8 仍利用第七题的资料，而把三个不同的地区视为因素 B 的三个水平，根据双向分类(无交互作用)多元方差分析考虑

① 对于 $\alpha=0.05$，检验各品种之间产量是否有显著差异，并与第七题结论加以比较。

② 对于 $\alpha=0.05$，检验不同地区的产量有否显著差异。

6.9 用调优运算法寻找塑料胶卷的最优工艺条件，考查三个指标：$x_1=$ 耐撕力，$x_2=$ 光泽，$x_3=$ 不透明度，考虑两个因素 $A=$ 拉速，$B=$ 添加剂的量，每个因素取两个水平，对每种因素组合做 5 次试验，其数据如下：

因素A \ 指标值 \ 因素B	指标值	B_1（低量）	B_2（高量）
A_1（低速）	$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$	$\begin{pmatrix}6.5\\9.5\\4.4\end{pmatrix},\begin{pmatrix}6.2\\9.9\\6.4\end{pmatrix},\begin{pmatrix}5.8\\9.6\\3.0\end{pmatrix},\begin{pmatrix}6.5\\9.6\\4.1\end{pmatrix},\begin{pmatrix}6.5\\9.2\\0.8\end{pmatrix}$	$\begin{pmatrix}6.9\\9.1\\5.7\end{pmatrix},\begin{pmatrix}7.2\\10.0\\2.0\end{pmatrix},\begin{pmatrix}6.9\\9.9\\3.9\end{pmatrix},\begin{pmatrix}6.1\\9.5\\1.9\end{pmatrix},\begin{pmatrix}6.3\\9.4\\5.7\end{pmatrix}$
A_2（高速）	$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$	$\begin{pmatrix}6.7\\9.1\\2.8\end{pmatrix},\begin{pmatrix}6.6\\9.3\\4.1\end{pmatrix},\begin{pmatrix}7.2\\8.3\\3.8\end{pmatrix},\begin{pmatrix}7.1\\8.4\\1.6\end{pmatrix},\begin{pmatrix}6.8\\8.5\\3.4\end{pmatrix}$	$\begin{pmatrix}7.1\\9.2\\8.4\end{pmatrix},\begin{pmatrix}7.0\\8.8\\5.2\end{pmatrix},\begin{pmatrix}7.2\\9.7\\6.9\end{pmatrix},\begin{pmatrix}7.5\\10.1\\2.7\end{pmatrix},\begin{pmatrix}7.6\\9.2\\1.9\end{pmatrix}$

给定 $\alpha=0.05$，试检验因素 A，B，及 $A\times B$ 对试验结果是否有显著影响。

第七章 判别分析

判别分析是根据所研究的个体的观测指标来推断该个体所属类型的一种统计方法. 在自然科学和社会科学的研究中经常会碰到这种统计问题. 例如在地质找矿中我们要根据某异常点的地质结构、化探和物探的各项指标来判断该异常点属于哪一种矿化类型; 医生要根据某人的各项化验指标的结果来判断该人属于什么病症;调查了某地区的土地生产率、劳动生产率、人均收入、费用水平, 农村工业比重等指标, 要确定该地区属于哪一种经济类型地区.

判别问题用统计的语言来表达,就是已有 q 个总体$\pi_1,\pi_2,\cdots,\pi_q$. 它们的分布函数分别为 $F_1(x)$, $F_2(x),\cdots,F_q(x)$, 每个 $F_i(x)$ 都是 p 维函数. 对于一个新给定的样品 x, 我们要判断它是来自哪一个总体. 当然, 我们应该要求判别准则在某种意义下是最优的, 例如错判的概率最小或错判的损失最小等. 本章仅介绍最基本的几种判别方法,即距离判别, Bayes 判别和 Fisher 判别.

§7.1 距离判别

(一) 统计距离

在判别分析中以及第八章的聚类分析中都要涉及到样品之间、样品与总体之间以及总体之间的距离. 因此我们对距离作一些简单介绍.

定义 1.1 设 $\mathscr{D}$ 是点集,$d(\cdot,\cdot)$ 是 $\mathscr{D}\times\mathscr{D}$ 到$[0,\infty)$的函数,对于 $\mathscr{D}$ 中的任意点 x_i,x_j 和 x_k, 若 $d(x_ix_j)$ (简记为 d_{ij})满足:

(i) $d_{ij}\geqslant 0$, 当且仅当 $x_i=x_j$ 时取等号;

(ii) $d_{ij}=d_{ji}$;

(iii) $d_{ij}\leqslant d_{ik}+d_{kj}$,(三角不等式)。

则称 d_{ij} 为点 x_i 与 x_j 的距离。

满足上述定义的 d_{ij} 有多种,下面列举一些常用的距离。

1°. 明考夫斯基距离(简称明氏距离)

$$d_{ij}(m)=\left[\sum_{\alpha=1}^{p}|x_{\alpha i}-x_{\alpha j}|^m\right]^{\frac{1}{m}},\ m>0 \tag{1.1}$$

当 $m=1,2,\infty$ 时分别得到绝对距离、欧氏距离和切比雪夫距离:

$$d_{ij}(1)=\sum_{\alpha=1}^{p}|x_{\alpha i}-x_{\alpha j}| \tag{1.2}$$

$$d_{ij}(2)=\left[\sum_{\alpha=1}^{p}(x_{\alpha i}-x_{\alpha j})^2\right]^{\frac{1}{2}} \tag{1.3}$$

$$d_{ij}(\infty)=\max_{1\leqslant\alpha\leqslant p}|x_{\alpha i}-x_{\alpha j}| \tag{1.4}$$

如果 x 的各个分量的测量值悬殊很大,采用明氏距离是不合理的,此时可以对数据先标准化或极差化处理后再来计算距离。

2°. B 模距离。

对于任给定的正定矩阵 B,由

$$d_{ij}=[(x_i-x_j)'B(x_i-x_j)]^{\frac{1}{2}} \tag{1.5}$$

算出的距离称为 x_i 与 x_j 的 B 模距离。

设任意点 x_i(随机向量)的 p 个分量的方差存在记为 σ_α^2,($\alpha=1,\cdots,p$) 而它们的数学期望向量和协方差阵分别为 $\mu=E(x_i)$, $\Sigma=\mathrm{cov}(x_i)$ ($\Sigma>0$) 时,那么当 $B=\Sigma^{-1}$ 时便得到了著名的 **Mahalanobis 距离**(简称马氏距离):

$$d_{ij}=[(x_i-x_j)'\Sigma^{-1}(x_i-x_j)]^{\frac{1}{2}} \tag{1.6}$$

读者可以验证马氏距离具有平移不变性和非奇异变换下的不变性。

当 $B=\mathrm{diag}\left(\frac{1}{\sigma_1^2},\cdots,\frac{1}{\sigma_p^2}\right)$ 时,便得到了方差加权距离:

$$d_{ij} = \left(\sum_{\alpha=1}^{p} \frac{(x_{\alpha i} - x_{\alpha j})^2}{\sigma_\alpha^2}\right)^{\frac{1}{2}} \tag{1.7}$$

当 $B = I$（单位阵）时，B 模距离便是(1.3)式的欧氏距离。

当点的各个分量对区分样品有不同的作用时，可给各分量不同的权重，例如第 α 个分量赋于权重 b_α，此时 B 可取为 $\mathrm{diag}(b_1, \cdots, b_p)$。我们可以粗略地说 B 的主对角线元素 $b_{\alpha\alpha}$ 表示第 α 个分量在区别样品时所占的权重，非对角线元素 $b_{\alpha\beta}$ 表示第 α 个分量与第 β 个分量的交互作用在区分样品时所占权重的一半。

（二） 两总体的情况

设有两个具有相同协方差阵 $\Sigma(\Sigma > 0)$ 的总体 π_1, π_2，它们均值向量分别为 $\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}$。对于一个新给定的样品 $\boldsymbol{x}$ 要判断它是来自于哪一个总体（或者说要判断它属于哪一个总体）。一个最直观的想法是分别计算 $\boldsymbol{x}$ 与两个总体的距离（点 $\boldsymbol{x}$ 到 $\boldsymbol{\mu}^{(i)}$ 的距离称为点 $\boldsymbol{x}$ 到总体 π_i 的距离）$d(\boldsymbol{x}, \boldsymbol{\mu}^{(i)})$，$i = 1, 2$。然后根据下列规则进行判别

$$\begin{cases} \boldsymbol{x} \in \pi_1 & \text{当 } d(\boldsymbol{x}, \boldsymbol{\mu}^{(1)}) \leqslant d(\boldsymbol{x}, \boldsymbol{\mu}^{(2)}) \\ \boldsymbol{x} \in \pi_2 & \text{当 } d(\boldsymbol{x}, \boldsymbol{\mu}^{(1)}) > d(\boldsymbol{x}, \boldsymbol{\mu}^{(2)}) \end{cases} \tag{1.8}$$

（当 $d(\boldsymbol{x}, \boldsymbol{\mu}^{(1)}) = d(\boldsymbol{x}, \boldsymbol{\mu}^{(2)})$ 时，$\boldsymbol{x}$ 可归属于 π_1, π_2 的任何一个，为了方便叙述不妨将它归属 π_1）。在这里我们采用马氏距离。为了简化(1.8)，计算两个马氏距离平方之差

$$\begin{aligned} d^2(\boldsymbol{x}, \boldsymbol{\mu}^{(1)}) - d^2(\boldsymbol{x}, \boldsymbol{\mu}^{(2)}) &= (\boldsymbol{x} - \boldsymbol{\mu}^{(1)})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}^{(1)}) \\ &\quad - (\boldsymbol{x} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}^{(2)}) \\ &= -2[\boldsymbol{x} - (\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})/2]'\Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) \\ &= -2W(\boldsymbol{x}) \end{aligned}$$

其中

$$W(\boldsymbol{x}) = (\boldsymbol{x} - \bar{\boldsymbol{\mu}})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) \tag{1.9}$$

$$\bar{\boldsymbol{\mu}} = \frac{1}{2}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})$$

于是在马氏距离之下规则(1.8)变为

$$\begin{cases} x \in \pi_1 & \text{当 } W(x) \geqslant 0 \\ x \in \pi_2 & \text{当 } W(x) < 0 \end{cases} \tag{1.10}$$

$W(x)$ 是 x 的一个线性函数，一般将 $W(x)$ 称为**线性判别函数**，显然 p 维平面 $W(x)=0$，它把 p 维空间分成两部分，即得到 p 维空间一个划分：

$$R_1 = \{x: W(x) \geqslant 0\}$$

$$R_2 = \{x: W(x) < 0\}$$

当样品 $x \in R_1$ 时，则判断 $x \in \pi_1$，当 $x \in R_2$，则判断 $x \in \pi_2$. 当我们引入总体先验概率、错判损失等概念后，如何寻找一个使平均错判损失达到最小的划分问题我们将在下一节的 Bayes 判别中来讨论。对于上述判别规则(1.10)作几点说明，它对于我们理解判别分析很重要.

1° 按最小距离规则判别是会产生误判的，为了说明问题，不妨设 $p=1$ 且 π_1 为 $N(\mu_1, \sigma^2)$，π_2 为 $N(\mu_2, \sigma^2)$ ($\mu_2 > \mu_1$) 由图 1.1 中可以看到：当 x 事实上取自 π_1，但它的观察值却落在 $\bar{\mu} = \frac{1}{2}(\mu_1 + \mu_2)$ 的右边，按上述的规则应把 x 判断为 π_2，此时发生误判，误判的概率正是图中的阴影部分的面积。另外**判别限**(或称**阈值**)的选取是很重要的，如果不以 $\bar{\mu}$ 为判别限，而以另一点 ξ 为判别限，这时将 π_1 误判为 π_2 的概率减小了，但 π_2 误判为 π_1 的概率却增大了。对于正态总体我们可直接验证最小距离判别的判别限 $\bar{\mu}$ 保证了这两个误判概率相同.



图 1.1

2° 从图 1.1 可看出，当两个总体 π_1,π_2 十分接近时，则无论用什么办法，误判概率都很大，这时判别是没有意义的，因此在判别之前应对两总体的均值是否有显著差异进行检验。

3° 由于落在 $\bar{\mu}$ 附近的点误判概率比较大，有时可划出一个待判区域，例如取 $[c,d]=\left[\bar{\mu}-\frac{1}{5}|\mu_1-\mu_2|,\bar{\mu}+\frac{1}{5}|\mu_1-\mu_2|\right]$ 作为待判区域，这样就有

$$\begin{cases}x\in\pi_1 & \text{当 } x\leqslant c\\ x\in\pi_2 & \text{当 } x\geqslant d\\ \text{待判} & \text{当 } c<x<d\end{cases}$$

4° 以上判别函数及规则并没有涉及到具体的分布类型，只要二阶矩存在就可以了。

如果两总体的均值向量及公共协方差阵未知时，可以用两总体的样本来估计。设 $x_i^{(1)}, i=1,2,\cdots,n_1$。$x_i^{(2)}, i=1,2,\cdots,n_2$ 分别是 π_1 和 π_2 的样本，令

$$\bar{x}^{(1)}=\frac{1}{n_1}\sum_{i=1}^{n_1}x_i^{(1)},\quad \bar{x}^{(2)}=\frac{1}{n_2}\sum_{i=1}^{n_2}x_i^{(2)}$$

$$A_1=\sum_{i=1}^{n_1}(x_i^{(1)}-\bar{x}^{(1)})(x_i^{(1)}-\bar{x}^{(1)})',$$

$$A_2=\sum_{i=1}^{n_2}(x_i^{(2)}-\bar{x}^{(2)})(x_i^{(2)}-\bar{x}^{(2)})'$$

$$\hat{\Sigma}=\frac{1}{n_1+n_2-2}(A_1+A_2)$$

那末判别函数可取为

$$W(x)=\left[x-\frac{1}{2}(\bar{x}^{(1)}+\bar{x}^{(2)})\right]'\hat{\Sigma}^{-1}(\bar{x}^{(1)}-\bar{x}^{(2)})\tag{1.11}$$

判别的规则与(1.10)一样。

上面讨论的是两总体具有相同的公共协方差阵的情况。如果两总体 π_1 与 π_2 的协方差阵 Σ_1 和 Σ_2 不相等，此时判别函数

$$W(x)=d^2(x,\mu^{(1)})-d^2(x,\mu^{(2)})$$
$$=(x-\mu^{(1)})'\Sigma_1^{-1}(x-\mu^{(1)})$$
$$-(x-\mu^{(2)})'\Sigma_2^{-1}(x-\mu^{(2)}) \qquad (1.12)$$

是 x 的二次函数，显然它比线性函数复杂。

（三） 多总体的情况

设有 q 个总体 $\pi_1,\cdots,\pi_q$。它们具有公共的正定协方差阵和不同的均值向量 $\mu^{(i)}$，$i=1,2,\cdots,q$。那么判别函数可取为（这里仍采用马氏距离）

$$W_{ij}(x)=\left[x-\frac{1}{2}(\mu^{(i)}+\mu^{(j)})\right]'\Sigma^{-1}(\mu^{(i)}-\mu^{(j)}),$$

$$i,j=1,2,\cdots,q \quad (1.13)$$

此时 p 维空间的划分 $\{R_1,\cdots,R_q\}$ 可如下规定：除了边界点以外，有

$$R_i=\{x:W_{ij}(x)>0,\ 对一切\ i\neq j\}\ i=1,2,\cdots,q \quad (1.14)$$

判别的规则为：若 $x\in R_i$，则判断 $x\in\pi_i$，$i=1,2,\cdots,q$。如果 x 使得 $W_{ij_1}(x)=W_{ij_2}(x)=\cdots=W_{ij_r}(x)=0$ $(1\leqslant r\leqslant q)$，则 x 可判属于 $\pi_i,\pi_{j_1},\cdots,\pi_{j_r}$ 的任一个，即在边界上的点可判断为相邻区域的任一个。

类似地，当 $\mu^{(i)}$，$i=1,2,\cdots,q$. 和 Σ 都是未知时，可用它们相应的估计代替。设 $x_\alpha^{(i)}$ 是 π_i 中抽取的容量分别为 n_i 的样本 $(i=1,2,\cdots,q.\ \alpha=1,2,\cdots,n_i)$。记

$$\bar{x}^{(i)}=\frac{1}{n_i}\sum_{\alpha=1}^{n_i}x_\alpha^{(i)},\quad A_i=\sum_{\alpha=1}^{n_i}(x_\alpha^{(i)}-\bar{x}^{(i)})(x_\alpha^{(i)}-\bar{x}^{(i)})'$$

$$\hat{\Sigma}=\frac{1}{n-q}\sum_{i=1}^{q}A_i,\quad 其中\ n=\sum_{i=1}^{q}n_i,$$
$$(i=1,2,\cdots,q)$$

相应的判别函数为

$$W_{ij}(x)=\left[x-\frac{1}{2}(\bar{x}^{(i)}+\bar{x}^{(j)})\right]'\hat{\Sigma}^{-1}(\bar{x}^{(i)}-\bar{x}^{(j)}) \quad (1.15)$$

判别的规则如前所述。

上面讨论的是 q 个总体有公共协方差阵 Σ，如果 q 个总体的协方差阵不同，设 π_i 的均值向量为 $\boldsymbol{\mu}^{(i)}$，协方差矩阵为 Σ_i ($\Sigma_i>0$)，$i=1,2,\cdots,q$. 此时判别区域 R_i, $i=1,2,\cdots,q$ 比较复杂，在实用上是直接去计算 $d^2(\boldsymbol{x},\boldsymbol{\mu}^{(i)})$，$i=1,2,\cdots,q$. 将 $\boldsymbol{x}$ 判入距离最小的哪一类，按此规则判别区域可写为

$$R_i=\{\boldsymbol{x}: d^2(\boldsymbol{x},\boldsymbol{\mu}^{(i)})=\min_{1\leqslant k\leqslant q} d^2(\boldsymbol{x},\boldsymbol{\mu}^{(k)})\}, i=1,2,\cdots,q \quad (1.16)$$

若 $\boldsymbol{x}$ 到 q 个总体的最小距离不止一个，即

$$d^2(\boldsymbol{x},\boldsymbol{\mu}^{(i_1)})=\cdots=d^2(\boldsymbol{x},\boldsymbol{\mu}^{(i_r)})=\min_{1\leqslant k\leqslant q} d^2(\boldsymbol{x},\boldsymbol{\mu}^{(k)}) \quad (1\leqslant r\leqslant q)$$

则 $\boldsymbol{x}$ 可归入 $\pi_{i_1},\cdots,\pi_{i_r}$ 的任何一个。

例 1.1 生物统计学家经常提出由个体的"大小"及"形状"因子(这就相当在第九章中将要介绍的两个主成份)来进行判别的问题，用 $\boldsymbol{x}=(x_1,x_2)'$ 表示"大小"和"形状"两个因子。现从正常人和精神病患者两个总体中各抽取了 25 个测量了某些指标并变换为"大小"和"形状"数据，列于表 1.1。另外还新测定了一"病员"的"大小"和"因子"数据列在该表的最后一行，试用最小距离法判别该"病员"是否患有精神病。

从表 1.1 中的数据计算得

$$\bar{\boldsymbol{x}}^{(1)}=\frac{1}{25}\sum_{\alpha=1}^{25}\boldsymbol{x}_\alpha^{(1)}=(20.80,\ 12.32)',$$

$$\bar{\boldsymbol{x}}^{(2)}=\frac{1}{25}\sum_{\alpha=1}^{25}\boldsymbol{x}_\alpha^{(2)}=(12.80,36.40)'$$

$$S_1=\frac{1}{24}A_1=\begin{pmatrix}6.90 & -5.27\\ -5.27 & 40.89\end{pmatrix},$$

$$S_2=\frac{1}{24}A_2=\begin{pmatrix}36.75 & 13.92\\ 13.92 & 287.92\end{pmatrix}$$

表 1.1　正常人和精神病患者的"大小"和"形状"数据

	正常人			精神病患者		
序号	$x_1^{(1)}$	$x_2^{(1)}$	$W^*(x)$	$x_1^{(2)}$	$x_2^{(2)}$	$W^*(x)$
1	22	6	62	24	38	8
2	20	14	36	19	36	−13
3	23	9	61	11	43	−67
4	23	1	77	6	60	−126
5	17	8	33	9	32	−55
6	24	9	66	10	17	−20
7	23	13	53	3	17	−55
8	18	18	18	15	56	−73
9	22	16	42	14	43	−52
10	19	18	23	20	8	48
11	20	17	30	8	46	−88
12	20	31	2	20	62	−60
13	21	9	51	14	36	−38
14	13	13	3	3	12	−45
15	20	14	36	10	51	−88
16	19	15	29	22	22	30
17	20	11	42	11	30	−41
18	18	17	20	6	30	−66
19	20	7	50	20	61	−58
20	23	6	67	20	43	−22
21	23	23	33	15	48	−57
22	25	9	71	5	53	−117
23	23	5	69	10	43	−72
24	21	12	45	13	19	−9
25	23	7	65	12	4	16
新病员	28	36	22			

$$\hat{\Sigma} = \frac{A_1 + A_2}{n_1 + n_2 - 2} = \frac{1}{2}\begin{pmatrix} 43.65 & 8.65 \\ 8.65 & 328.81 \end{pmatrix}$$

$$= \begin{pmatrix} 21.83 & 4.33 \\ 4.33 & 164.40 \end{pmatrix},$$

$$\bar{x}^{(1)} - \bar{x}^{(2)} = (8.00,\ -24.08)'$$

取 $\hat{\Sigma}$ 的伴随矩阵

$$V = \begin{pmatrix} 164.40 & -4.33 \\ -4.33 & 21.83 \end{pmatrix}$$

则

$$\hat{\Sigma} \cdot V = \begin{pmatrix} 3570.1 & 0 \\ 0 & 3570.1 \end{pmatrix} = 3570.1I \quad (I \text{ 为单位矩阵})$$

从而

$$\hat{\Sigma}^{-1} = \frac{1}{3570.1}\begin{pmatrix} 164.40 & -4.33 \\ -4.33 & 21.83 \end{pmatrix}$$

将上述计算得的数据代入线性判别函数(1.9)得

$$W(x) = \frac{1}{3570.1}(1419x_1 - 560x_2 - 10198)$$

$$\approx \frac{280}{3570.1}(5x_1 - 2x_2 - 36)$$

$$= \frac{280}{3570.1}W^*(x)$$

显然 $W(x) \geqslant 0$ 或者 $W(x) < 0$ 等价于 $W^*(x) = 5x_1 - 2x_2 - 36 \geqslant 0$ 或者 $W^*(x) < 0$. 于是，我们按 $W^*(x) \geqslant 0$ 判为正常人，$W^*(x) < 0$ 判为精神病患者. 当我们用原数据进行回判时，25 个正常人的数据代入 $W^*(x)$ 后全部大于 0，而 25 个精神病患者代入后有 4 个大于 0。误判率为 8%. 把新测定的“病员”数据代入后得 $W^*(x) = 22 > 0$，故该“病员”应判为正常人。

§7.2 Bayes 判 别

(一) 误判概率与误判损失

关于误判概率的概念在上一节中已谈过. 设有两个总体 π_1 和 π_2，根据某一个判别规则，将实际上为 π_1 的个体判为 π_2 或者将实际上为 π_2 的个体判为 π_1 的概率就是误判概率，一个好的判别规则

应该是使误判概率最小。除此之外还有一个误判损失问题或者说误判产生的花费（cost）问题，如把 π_1 的个体误判到 π_2 的损失比 π_2 的个体误判到 π_1 严重得多，则人们在作前一种判断时就要特别谨慎。譬如在药品检验中把有毒的个体判为无毒的其后果比无毒个体判为有毒的严重得多，因此一个好的判别规则还必须使得误判损失最小。

为了说明问题，我们仍以两个总体的情况来讨论。设所考虑的两个总体 π_1 与 π_2 分别具有密度函数 $f_1(\boldsymbol{x})$ 与 $f_2(\boldsymbol{x})$，其中 $\boldsymbol{x}$ 为 p 维向量。记 Ω 为所有可能观测值 $\boldsymbol{x}$ 的全体，称它为样本空间，R_1 为根据我们的规则要判为 π_1 的那些 $\boldsymbol{x}$ 的全体，而 $R_2=\Omega-R_1$ 是要判为 π_2 的那些 $\boldsymbol{x}$ 的全体。显然 R_1 与 R_2 互斥完备。某个体实际是来自 π_1，但被判为 π_2 的概率为

$$P(2/1)=P(\boldsymbol{x}\in R_2/\pi_1)=\int\cdots\int_{R_2} f_1(\boldsymbol{x})d\boldsymbol{x} \qquad (2.1)$$

来自 π_1 被判为 π_1 的概率为

$$P(1/1)=P(\boldsymbol{x}\in R_1/\pi_1)=\int\cdots\int_{R_1} f_1(\boldsymbol{x})d\boldsymbol{x} \qquad (2.2)$$

类似地，来自 π_2 被判为 π_1 以及来自 π_2 被判为 π_2 的概率分别为

$$P(1/2)=P(\boldsymbol{x}\in R_1/\pi_2)=\int\cdots\int_{R_1} f_2(\boldsymbol{x})d\boldsymbol{x} \qquad (2.3)$$

$$P(2/2)=P(\boldsymbol{x}\in R_2/\pi_2)=\int\cdots\int_{R_2} f_2(\boldsymbol{x})d\boldsymbol{x} \qquad (2.4)$$

又设 p_1 和 p_2 分别表示总体 π_1 和 π_2 的先验概率，且 $p_1+p_2=1$。于是

$$\begin{aligned} P\,(\text{正确地判为 }\pi_1) &= P\,(\text{来自 }\pi_1\text{，被判为 }\pi_1) \\ &= P(\boldsymbol{x}\in R_1/\pi_1)\cdot P(\pi_1)=P(1/1)\cdot p_1 \end{aligned} \qquad (2.5)$$

$$\begin{aligned} P\,(\text{误判到 }\pi_1) &= P\,(\text{来自 }\pi_2\text{，但判为 }\pi_1) \\ &= P(\boldsymbol{x}\in R_1/\pi_2)\cdot P(\pi_2)=P(1/2)\cdot p_2 \end{aligned} \qquad (2.6)$$

类似地有

$$P\ (\text{正确地判为}\ \pi_2) = P(2/2)\cdot p_2 \tag{2.7}$$

$$P\ (\text{误判到}\ \pi_2) = P(2/1)p_1 \tag{2.8}$$

设 $C(1/2)$ 表示来自 π_2 误判为 π_1 引起的损失，$C(2/1)$ 表示来自 π_1 误判为 π_2 引起的损失，并规定 $C(1/1) = C(2/2) = 0$.

将上述的误判概率与误判损失结合起来定义一个**平均误判损失**（expected cost of misclassification）并记为 ECM.

$$\mathrm{ECM}(R_1, R_2) = C(2/1)P(2/1)p_1 + C(1/2)P(1/2)p_2 \tag{2.9}$$

一个合理的判别规则应使 ECM 达到极小。

对于(2.9)式可以推广到 q 个总体的情况，设 q 个总体 π_i，$i = 1,2,\cdots,q$ 的密度函数为 $f_i(\boldsymbol{x})$，先验概率为 p_i，$i=1,2,\cdots,q$. 用 $C(k/i)$ 表示实际上来自 π_i 的个体误判到 π_k 产生的损失 $(i, k = 1, 2, \cdots, q)$. 约定 $C(i/i) = 0$. 又设 R_k 是根据规则判为 π_k 的那些 $\boldsymbol{x}$ 组成的集合，且 R_k 互斥完备. $P(k/i) = P(\boldsymbol{x} \in R_k/\pi_i) = \int\cdots\int_{R_k} f_i(\boldsymbol{x})d\boldsymbol{x}$，$(i, k = 1, 2, \cdots, q.)$ 且 $P(i/i) = 1 - \sum_{\substack{k=1\\k\neq i}}^{q} P(k/i)$.

来自 π_i 的个体被误判的平均损失为

$$\mathrm{ECM}(R_i) = \sum_{\substack{k=1\\k\neq i}}^{q} P(k/i)C(k/i) \quad i = 1,2,\cdots,q$$

于是总的平均误判损失为

$$\mathrm{ECM}(R_1,\cdots,R_q) = \sum_{i=1}^{q}\left[p_i \sum_{\substack{k=1\\k\neq i}}^{q} P(k/i)C(k/i)\right] \tag{2.10}$$

(2.10)式便是(2.9)式在多总体情况下的推广。因此，对于多总体情况一个合理的判别规则应使(2.10)达到极小。

（二） 两总体的 Bayes 判别规则

由（一）中的叙述知道，我们要选择样本空间 Ω 的一个划分：R_1 和 $R_2 = \Omega - R_1$ 使得平均损失(2.9)式达到极小。

定理 2.1 极小化平均损失(2.9)的区域 R_1 和 R_2 为:

$$R_1 = \left\{x: \frac{f_1(x)}{f_2(x)} \geqslant \frac{C(1/2)}{C(2/1)} \cdot \frac{p_2}{p_1}\right\}. \tag{2.11}$$

$$R_2 = \left\{x: \frac{f_1(x)}{f_2(x)} < \frac{C(1/2)}{C(2/1)} \cdot \frac{p_2}{p_1}\right\} \tag{2.12}$$

(当 $f_1(x)/f_2(x) = C(1/2) \cdot p_2/C(2/1) \cdot p_1$ 时，即 x 为边界点，它可归入 R_1, R_2 的任何一个，为了方便就将它归入 R_1)

证明 将 $P(2/1) = \int \cdots \int_{R_2} f_1(x)dx$, 及

$$P(1/2) = \int \cdots \int_{R_1} f_2(x)dx$$

代入(2.9)式得

$$\mathrm{ECM}(R_1, R_2) = C(2/1) \cdot p_1 \int \cdots \int_{R_2} f_1(x)dx + C(1/2) \cdot p_2 \int \cdots \int_{R_1} f_2(x)dx$$

注意到

$$\int \cdots \int_{\Omega} f_1(x)dx = \int \cdots \int_{R_1} f_1(x)dx + \int \cdots \int_{R_2} f_1(x)dx = 1,$$

有

$$\mathrm{ECM}(R_1, R_2) = \int \cdots \int_{R_1} [C(1/2)p_2f_2(x) - C(2/1)p_1f_1(x)]dx + C(2/1)p_1$$

要使 ECM 达到极小，仅当 R_1 中包含的点 x 满足不等式

$$C(1/2)p_2f_2(x) - C(2/1)p_1f_1(x) \leqslant 0$$

这样便证明了(2.11)式。同样也可证明(2.12)式。▮

由上述定理我们便得到两总体的 Bayes 判别规则:

$$
\begin{cases} x \in \pi_1 & \text{当 } x \text{ 使得 } \dfrac{f_1(x)}{f_2(x)} \geqslant \dfrac{C(1/2)}{C(2/1)} \cdot \dfrac{p_2}{p_1} \\ x \in \pi_2 & \text{当 } x \text{ 使得 } \dfrac{f_1(x)}{f_2(x)} < \dfrac{C(1/2)}{C(2/1)} \cdot \dfrac{p_2}{p_1} \end{cases} \tag{2.13}
$$

应用此规则时仅仅需要计算:

i) 新样品点 $x_0=(x_{01},\cdots,x_{0p})'$ 的密度函数比 $f_1(x_0)/f_2(x_0)$;

ii) 损失比 $C(1/2)/C(2/1)$;

iii) 先验概率比 p_2/p_1。

损失和先验概率以比值的形式出现是很重要的，因为确定两种损失的比值(或两总体的先验概率的比值)往往比确定损失本身(或先验概率本身)来得容易。下面列举(2.13)的三种特殊情况

i) 当 $p_2/p_1 = 1$ 时

$$
\begin{cases} x_1 \in \pi_1 & \text{当 } x \text{ 使得 } \dfrac{f_1(x)}{f_2(x)} \geqslant \dfrac{C(1/2)}{C(2/1)} \\ x_2 \in \pi_2 & \text{当 } x \text{ 使得 } \dfrac{f_1(x)}{f_2(x)} < \dfrac{C(1/2)}{C(2/1)} \end{cases} \tag{2.14}
$$

ii) 当 $C(1/2)/C(2/1) = 1$ 时

$$
\begin{cases} x \in \pi_1 & \text{当 } x \text{ 使得 } f_1(x)/f_2(x) \geqslant p_2/p_1 \\ x \in \pi_2 & \text{当 } x \text{ 使得 } f_1(x)/f_2(x) < p_2/p_1 \end{cases} \tag{2.15}
$$

iii) 当 $p_2/p_1 = C(1/2)/C(2/1) = 1$ 时

$$
\begin{cases} x \in \pi_1 & \text{当 } x \text{ 使得 } f_1(x)/f_2(x) \geqslant 1 \\ x \in \pi_2 & \text{当 } x \text{ 使得 } f_1(x)/f_2(x) < 1 \end{cases} \tag{2.16}
$$

对于具体问题，如果先验概率或者其比值都是难于确定，此时就利用规则(2.14)，同样如误判损失或者其比值都是难以确定，此时就利用规则(2.15)，如果上述两者都难以确定则利用规则(2.16)，最后这种情况是一种无可奈何的办法，当然判别也变得很简单：若 $f_1(x) \geqslant f_2(x)$ 则判 $x \in \pi_1$，否则判 $x \in \pi_2$。

我们将上述的两总体 Bayes 判别应用于正态总体 π_i: $N_p(\mu^{(i)},\Sigma_i)$ $(i = 1,2)$，分两种情况讨论。

$1°.\quad \Sigma_1 = \Sigma_2(=\Sigma)(\Sigma > 0)$

此时 π_i 的密度函数为

$$f_i(\boldsymbol{x}) = (2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})\right\} \tag{2.17}$$

定理 2.2 设总体 π_i 为 $N_p(\boldsymbol{\mu}^{(i)}, \Sigma)$ $(i=1,2)$ 其中 $\Sigma > 0$ 则使得平均误判损失极小的划分为

$$\begin{cases} R_1 = \{\boldsymbol{x}: W(\boldsymbol{x}) \geqslant \beta\} & (2.18) \\ R_2 = \{\boldsymbol{x}: W(\boldsymbol{x}) < \beta\} & (2.19) \end{cases}$$

其中

$$W(\boldsymbol{x}) = \left[\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})\right]'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \tag{2.20}$$

$$\beta = \ln\frac{C(1/2)\cdot p_2}{C(2/1)\cdot p_1} \tag{2.21}$$

证明：将密度函数(2.17)分别代入(2.11)与(2.12)得

$$\begin{cases} R_1 = \left\{\boldsymbol{x}: \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})\right] \geqslant \frac{C(1/2)p_2}{C(2/1)p_1}\right\} \\ R_2 = \left\{\boldsymbol{x}: \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})\right] < \frac{C(1/2)p_2}{C(2/1)p_1}\right\} \end{cases}$$

注意到指数函数的单调性以及等式

$$-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)}) = \left[\boldsymbol{x}-\frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})\right]'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})$$

这样便得到(2.18)和(2.19) ▋

不难发现 (2.20) 式的 $W(\boldsymbol{x})$ 与上节最小距离判别的线性判别函数(1.9)是一致的. 判别规则也仅只是判别限不一样，一个是 β，一个是 0. 如果当 $C(1/2)p_2/(C(2/1)p_1)=1$ 时，两者就完全一致。

如果总体的 $\boldsymbol{\mu}^{(1)}$、$\boldsymbol{\mu}^{(2)}$ 及 Σ 未知，与以往一样. 用总体的样本算出 $\bar{\boldsymbol{x}}^{(1)}$、$\bar{\boldsymbol{x}}^{(2)}$ 和 $S=\dfrac{1}{n_1+n_2-2}(A_1+A_2)$ 来代替 $\boldsymbol{\mu}^{(1)}$、$\boldsymbol{\mu}^{(2)}$ 和 Σ. 这样判别函数(2.20)可写为

$$W(\boldsymbol{x})=\left[\boldsymbol{x}-\frac{1}{2}(\bar{\boldsymbol{x}}^{(1)}+\bar{\boldsymbol{x}}^{(2)})\right]' S^{-1}(\bar{\boldsymbol{x}}^{(1)}-\bar{\boldsymbol{x}}^{(2)}) \quad (2.22)$$

我们把(2.22)称为 Anderson 线性判别函数. 判别的规则为

$$\begin{cases}\boldsymbol{x}\in\pi_1 & \text{当 } \boldsymbol{x} \text{ 使得 } W(\boldsymbol{x})\geqslant\beta \\ \boldsymbol{x}\in\pi_2 & \text{当 } \boldsymbol{x} \text{ 使得 } W(\boldsymbol{x})<\beta\end{cases} \quad (2.23)$$

其中 β 为(2.21)所规定.

这里应该指出，总体参数用其估计来代替，所得到的规则，仅仅只是最优（在平均误判损失达到极小的意义下）规则的一个估计，这时对于一个具体问题来讲，我们并没有把握说所得到的规则能够使平均误判损失达到最小，但当样本的容量充分大时，估计 $\bar{\boldsymbol{x}}^{(1)}$, $\bar{\boldsymbol{x}}^{(2)}$, S 分别和 $\boldsymbol{\mu}^{(1)}$、$\boldsymbol{\mu}^{(2)}$、Σ 很接近，因此我们有理由认为“样本”判别规则的性质会很好的.

2° $\Sigma_1\neq\Sigma_2$ ($\Sigma_1>0$, $\Sigma_2>0$).

由于误判损失极小化的划分依赖于密度函数之比 $f_1(\boldsymbol{x})/f_2(\boldsymbol{x})$ 或等价于它的对数 $\ln(f_1(\boldsymbol{x})/f_2(\boldsymbol{x}))$, 把协方差矩阵不等的两个多元正态密度代入这个比后，包含 $|\Sigma_i|^{\frac{1}{2}}$, $(i=1,2)$ 的因子不能消去，而且 $f_i(\boldsymbol{x})$ 的指数部分也不能组合成简单表达式，因此对于 $\Sigma_1\neq\Sigma_2$ 时，由(2.11)和(2.12)可得判别区域:

$$\begin{cases}R_1=\{\boldsymbol{x}:W^*(\boldsymbol{x})\geqslant K\} \\ R_2=\{\boldsymbol{x}:W^*(\boldsymbol{x})<K\}\end{cases}$$

其中

$$W^*(x) = -\frac{1}{2}x'(\Sigma_1^{-1} - \Sigma_2^{-1})x + (\mu^{(1)'}\Sigma_1^{-1} - \mu^{(2)'}\Sigma_2^{-1})x \tag{2.24}$$

$$K = \ln\left(\frac{C(1/2)p_2}{C(2/1)p_1}\right) + \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}(\mu^{(1)'}\Sigma_1^{-1}\mu^{(1)} - \mu^{(2)'}\Sigma_2^{-1}\mu^{(2)}) \tag{2.25}$$

显然，判别函数 $W^*(x)$ 是二次函数，它比 $\Sigma_1 = \Sigma_2$ 时的情况复杂得多。如果 $\mu^{(i)}$，$\Sigma_i(i = 1,2)$ 未知，仍可采用其估计来代替。

例 3.1 表 2.1 是某气象站预报有无春旱的实际资料，x_1 与 x_2 都是综合预报因子（气象含义从略），有春旱的是 6 个年份的资料，无春旱的是 8 个年份的资料，它们的先验概率分别用 $\frac{6}{14}$ 和 $\frac{8}{14}$ 来估计，并设误判损失相等，试建立 Anderson 线性判别函数。

表 2.1 某气象站预报有无春旱的资料

	序 号	1	2	3	4	5	6	7	8
春旱	x_1	24.8	24.1	26.6	23.5	25.5	27.4		
	x_2	−2.0	−2.4	−3.0	−1.9	−2.1	−3.1		
	$W(x_1,x_2)$	6.886	6.907	7.790	6.527*	7.100	8.029		
无春旱	x_1	22.1	21.6	22.0	22.8	22.7	21.5	22.1	21.4
	x_2	−0.7	−1.4	−0.8	−1.6	−1.5	−1.0	−1.2	−1.3
	$W(x_1,x_2)$	5.624	5.835	5.647	6.217	6.146	5.622	5.861	5.740

由表 2.1 中的数据计算得

$$\bar{x}^{(1)} = (25.32, -2.42)', \quad \bar{x}^{(2)} = (22.03, -1.19)'$$

$$\bar{x}^{(1)} - \bar{x}^{(2)} = (3.29, -1.23)',$$

$$\frac{1}{2}(\bar{x}^{(1)}+\bar{x}^{(2)})=(23.68,\ -1.81)'$$

$$S_1=\begin{pmatrix}2.220 & -0.640\\ -0.640 & 0.270\end{pmatrix},\quad S_2=\begin{pmatrix}0.270 & 0.007\\ 0.007 & 0.1070\end{pmatrix}$$

$$S=\frac{1}{12}(5S_1+7S_2)=\begin{pmatrix}1.08 & -0.26\\ -0.26 & 0.17\end{pmatrix}$$

$$S^{-1}=\frac{1}{0.116}\begin{pmatrix}0.17 & 0.26\\ 0.26 & 1.08\end{pmatrix},$$

$$\beta=\ln\frac{p_2}{p_1}=\ln\frac{8}{6}=0.288$$

将上述计算结果代入(2.22)式得 Anderson 线性判别函数:

$$\frac{1}{0.116}(0.2395x_1-0.4730x_2-6.527)$$

为了计算和验证方便取判别函数为

$$W(x)=W(x_1,x_2)=0.2395x_1-0.4730x_2$$

取判别限为

$$6.527+\frac{1}{0.116}\times 0.288=6.560$$

将表 2.1 的数据代入 $W(x)$ 进行回报，计算的结果填在表 2.1 中 $W(x_1,x_2)$ 相应的栏目中，回报中错报的只有一个，即春旱中的第 4 号，与历史资料的拟合率达 93%。

(三) 多个总体的 Bayes 判别规则

在(一)的讨论中已知道，对于 q 个总体的情况平均误判损失为(2.10)式，即

$$\mathrm{ECM}(R_1,\cdots,R_q)=\sum_{i=1}^{q}\left[p_i\sum_{\substack{k=1\\k\neq i}}^{q}P(k/i)C(k/i)\right]$$

我们要寻找一个互斥完备的区域 $R_1,\cdots,R_q$ 使它达到最小。令

$$h_k(x)=\sum_{\substack{i=1\\i\neq k}}^{q}p_if_i(x)\cdot C(k/i) \tag{2.26}$$

其中 $f_i(x)$ 为第 i 个总体 π_i 的密度函数。对于满足

$$h_j(x)=\min_{1\leqslant k\leqslant q} h_k(x) \tag{2.27}$$

的样本空间 Ω 中的样品点 x 组成的集合记为 R_j，如果有某个样品点 x 有 $r(1<r\leqslant q)$ 个标号 $j_1,\cdots,j_r$ 使得(2.27)式成立，即

$$h_{j_1}(x)=\cdots=h_{j_r}(x)=\min_{1\leqslant k\leqslant q} h_k(x) \tag{2.28}$$

那么我们规定该 x 可归入集合 $R_{j_1},\cdots,R_{j_r}$ 中的任何一个。这样就可将 Ω 划分为互斥完备的区域 $R_1,\cdots,R_q$。称此划分为 Bayes 划分。

定理 2.3 Bayes 划分 $R_1,\cdots,R_q$ 使得平均误判损失(2.10)达到极小。

证明 由于 $P(j/i)=\int\cdots\int_{R_j} f_i(x)dx$，再根据(2.10)及(2.26)式有

$$\begin{aligned}\mathrm{ECM}(R_1,\cdots,R_q)&=\sum_{i=1}^{q}p_i\sum_{\substack{k=1\\k\neq i}}^{q}P(k/i)C(k/i)\\&=\sum_{i=1}^{q}p_i\sum_{\substack{k=1\\k\neq i}}^{q}C(k/i)\int\cdots\int_{R_k}f_i(x)dx\\&=\sum_{k=1}^{q}\int\cdots\int_{R_k}h_k(x)dx\end{aligned}$$

如果 $R_1^*,\cdots,R_q^*$ 是 Ω 的另一个划分，其平均误判损失为

$$\mathrm{ECM}(R_1^*,\cdots,R_q^*)=\sum_{k=1}^{q}\int\cdots\int_{R_k^*}h_k(x)dx$$

于是

$$\begin{aligned}&\mathrm{ECM}(R_1,\cdots,R_q)-\mathrm{ECM}(R_1^*,\cdots,R_q^*)\\&=\sum_{i=1}^{q}\sum_{j=1}^{q}\int\cdots\int_{R_j\cap R_i^*}(h_i(x)-h_j(x))dx\end{aligned}$$

由于在 R_j 中的点 $\boldsymbol{x}$ 是满足(2.27)式，因此在 $R_i \cap R_j^*$ 中的 $\boldsymbol{x}$ 使得 $(h_j(\boldsymbol{x}) - h_i(\boldsymbol{x})) \leqslant 0$，故上式小于或等于 0，从而 Bayes 划分 $R_1, \cdots, R_q$ 使得平均误判损失最小。 ▌

根据定理 2.3 便可得到多总体的 Bayes 判别规则：

$$\text{当 } \boldsymbol{x} \text{ 使得 } \min_{1 \leqslant i \leqslant q} h_i(\boldsymbol{x}) = h_j(\boldsymbol{x}) \text{ 时，判 } \boldsymbol{x} \in \pi_j \tag{2.29}$$

(如 $\boldsymbol{x}$ 使得 $\min_{1 \leqslant i \leqslant q} h_i(\boldsymbol{x}) = h_{j_1}(\boldsymbol{x}) = \cdots = h_{j_r}(\boldsymbol{x})$，则判 $\boldsymbol{x}$ 为 $\pi_{j_1}, \cdots, \pi_{j_r}$ 中的任何一个)。如果误判损失都相同，不妨假定都为 1，此时 $h_k(\boldsymbol{x}) = \sum_{\substack{i=1 \\ i \neq k}}^{q} p_i f_i(\boldsymbol{x})$，$(k = 1, 2, \cdots, q)$，并注意到

$h_k(\boldsymbol{x}) = \sum_{i=1}^{q} p_i f_i(\boldsymbol{x}) - p_k f_k(\boldsymbol{x})$，且 $p_k \cdot f_k(\boldsymbol{x})$ 最大等价于 $h_k(\boldsymbol{x})$ 最小。于是对于误判损失相等的情况，有如下简单的判别规则：

当 $\boldsymbol{x}$ 使得 $p_j f_j(\boldsymbol{x}) > p_i f_i(\boldsymbol{x})$，$(j \neq i,\ i = 1, \cdots, q)$ 时，判 $\boldsymbol{x} \in \pi_j$ (2.30)

或者

当 $\boldsymbol{x}$ 使得 $\ln[p_j f_j(\boldsymbol{x})] > \ln[p_i f_i(\boldsymbol{x})]$ 对所有 $i \neq j$，判 $\boldsymbol{x} \in \pi_j$. (2.31)

规则(2.30)与后验概率最大规则是相同的。

一般 Bayes 规则（ECM 极小化法则）包含了三类量：先验概率、误判损失和密度函数。因此把这个规则付诸实现时，必须定出前两类量的值，并利用样本对密度函数中的未知参数作出估计。

我们将上述多总体的 Bayes 判别规则应用到正态总体，且只讨论误判损失相等的情况。

设 $\pi_i, i = 1, 2, \cdots, q$. 都是正态总体 $N_p(\boldsymbol{\mu}^{(i)}, \Sigma_i)$，$\Sigma_i > 0$. 其密度为

$$f_i(\boldsymbol{x}) = (2\pi)^{-\frac{p}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}^{(i)})' \Sigma_i^{-1} (\boldsymbol{x} - \boldsymbol{\mu}^{(i)})\right\},$$

$$i = 1, \cdots, q.$$

假定误判损失都相等，判别规则(2.31)成为

$$\text{当} \quad \ln[p_j f_j(x)] = \ln p_j - \frac{p}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| - \frac{1}{2}(x-\mu^{(j)})'\Sigma_j^{-1}(x-\mu^{(j)})$$

$$= \max_{1\leq k\leq q}\left\{\ln p_k - \frac{p}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(x-\mu^{(k)})'\Sigma_k^{-1}(x-\mu^{(k)})\right\} \text{时}$$

则判 x 归给 π_j. 消去与 k, j 无关的常数项 $-\frac{p}{2}\ln 2\pi$, 并记

$$d_k(x) = \ln p_k - \frac{1}{2}[\ln|\Sigma_k| + (x-\mu^{(k)})'\Sigma_k^{-1}(x-\mu^{(k)})]$$

$$k = 1, 2, \cdots q \qquad (2.32)$$

称 $d_k(x)$ 为样品 x 在第 k 个总体的二次判别得分. $d_k(x)$ 综合了广义方差 $|\Sigma_k|$, 先验概率 p_k 以及 x 到 $\mu^{(k)}$ 的马氏距离对判别的"贡献". 此时判别规则可简洁地写为: 若样品 x 使得

$$d_j(x) = \max_{1\leq k\leq q}\{d_k(x)\} \qquad (2.33)$$

则 x 判给总体 π_j.

在实际应用中，若诸 $\mu^{(k)}$, Σ_k 未知，则用样本对它们作出估计.

如果诸总体的协方差阵相等，即 $\Sigma_1 = \cdots = \Sigma_q (= \Sigma)$, (2.32)式变成了(消去了与 k 无关的项)

$$d_k^*(x) = x'\Sigma^{-1}\mu^{(k)} - \frac{1}{2}\mu^{(k)'}\Sigma^{-1}\mu^{(k)} + \ln p_k \qquad (2.34)$$

显然, $d_k^*(x)$ 是线性的，判别的规则与(2.33)相同，即样品 x 使得

$$d_j^*(x) = \max_{1\leq k\leq q}\{d_k^*(x)\} \qquad (2.35)$$

则 x 判给 π_j. 若 $\mu^{(k)}$ 与 Σ 未知，那么用样本对它们作出估计，注

意到，此时 Σ 的估计是

$$\hat{\Sigma}=S=\frac{(n_1-1)S_1+\cdots+(n_q-1)S_q}{(n_1+\cdots+n_q-q)} \tag{2.36}$$

其中 n_k,S_k 分别为总体 π_k 的样本容量及样本协方差阵。

例 2.2 对三种鸢尾花分别抽取了 $n_1=n_2=n_3=50$ 的样本，并对两个变量 $x_1=$ 花瓣宽，$x_2=$ 花萼宽进行测量，计算得

$$\pi_1:\bar{x}^{(1)}=(3.46,\ 0.25)' \quad S_1=\begin{pmatrix}0.0661 & 0.0298\\ 0.0298 & 0.0601\end{pmatrix}$$

$$\pi_2:\bar{x}^{(2)}=(2.77,\ 1.30)' \quad S_2=\begin{pmatrix}0.0711 & 0.0333\\ 0.0333 & 0.0400\end{pmatrix}$$

$$\pi_3:\bar{x}^{(3)}=(2.79,\ 2.01)' \quad S_3=\begin{pmatrix}0.0588 & 0.0323\\ 0.0323 & 0.0810\end{pmatrix}$$

三种鸢尾花的两个变量服从二维正态分布且有相同的协方差阵 Σ（参看第五章例 5.2），由 S_1,S_2 和 S_3 对 Σ 的估计为

$$S=\frac{49}{147}(S_1+S_2+S_3)=\begin{pmatrix}0.0653 & 0.0318\\ 0.0318 & 0.0604\end{pmatrix}$$

$$S^{-1}=\frac{1}{0.0029}\begin{pmatrix}0.0604 & -0.0318\\ -0.0318 & 0.0653\end{pmatrix}$$

设三个总体的先验概率相等以及误判损失相同，消去与标号无关的因子与项，三个线性判别得分为

$$d_1^*(x)=0.2009x_1-0.0937x_2-0.3358$$
$$d_2^*(x)=0.1259x_1+0.0032x_2-0.1723$$
$$d_3^*(x)=0.1045x_1+0.0425x_2-0.1885$$

对于新给的样品 $x_0=(2.8,1.5)'$ 应判给哪一个总体？将 x_0 代入 $d_k^*(x)$ 得

$$d_1^*(x_0)=0.0861,\ d_2^*(x_0)=0.1754,\ d_3^*(x_0)=0.1679$$

最大者为 $d_2^*(x_0)$，根据判别法则知 x_0 应判给 π_2 即判为第二种鸢尾花。

（四） 判别规则的评价

判别规则性能的评价，主要是通过计算它的误判概率的大小来衡量．对于两个总体的情况，由(2.6)和(2.8)知道对于某判别规则所确定的判别区域 R_1，R_2 产生的**误判概率总和**（total probability of misclassification）为

$$TPM = p_1 \int \cdots \int_{R_2} f_1(\boldsymbol{x}) d\boldsymbol{x} + p_2 \int \cdots \int_{R_1} f_2(\boldsymbol{x}) d\boldsymbol{x} \qquad (2.37)$$

因此计算某判别规则的总误判概率时，必须知道总体的密度、先验概率及判别区域．

例 2.3　对二个正态总体 $N_p(\boldsymbol{\mu}^{(i)}, \Sigma)$ $i=1,2$，并假定 $p_1 = p_2 = \frac{1}{2}$，误判损失相等，且设为 1，求 Bayes 判别的总误判概率．

由定理 2.2 知道 Bayes 的判别区域为

$$\begin{cases} R_1 = \{\boldsymbol{x}: W(\boldsymbol{x}) \geq 0\} \\ R_2 = \{\boldsymbol{x}: W(\boldsymbol{x}) < 0\} \end{cases}$$

其中 $W(\boldsymbol{x}) = \left[\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})\right]' \Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$，由于 $\boldsymbol{x}$ 为 p 元正态分布，根据多元正态分布的性质知道 $W(\boldsymbol{x})$ 为一元正态，其均值为

$$\nu_1 = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})' \Sigma^{-1} \boldsymbol{\mu}^{(1)} - \frac{1}{2}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})' \Sigma^{-1}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})$$

$$= \frac{1}{2}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})' \Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) = \frac{\Delta^2}{2} \quad \text{当 } \boldsymbol{x} \in \pi_1$$

$$\nu_2 = -\frac{1}{2}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})' \Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) = -\frac{\Delta^2}{2} \quad \text{当 } \boldsymbol{x} \in \pi_2$$

其方差为

$\sigma_W^2 = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})' \Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) = \Delta^2$ 无论 $\boldsymbol{x} \in \pi_1$ 或 $\boldsymbol{x} \in \pi_2$

上面式子中的 $\Delta^2 = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})' \Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$，它正好是 π_1 与 π_2 的马氏距离平方．(2.37)式的总误判概率可写为

$$\text{TPM}=\frac{1}{2}[P(2/1)+P(1/2)]$$

而

$$\begin{aligned}P(2/1)&=P(W(\boldsymbol{x})<0\mid\boldsymbol{x}\in\pi_1)\\&=P(W(\boldsymbol{x})<0\mid W(\boldsymbol{x})\sim N(\nu_1,\Delta^2))\\&=P\left(\frac{W(\boldsymbol{x})-\nu_1}{\Delta}<\frac{-\nu_1}{\Delta}\middle|\boldsymbol{x}\in\pi_1\right)\\&=P\left(\frac{W(\boldsymbol{x})-\nu_1}{\Delta}<-\frac{\Delta}{2}\middle|\boldsymbol{x}\in\pi_1\right)\\&=\Phi\left(-\frac{\Delta}{2}\right)\end{aligned}$$

类似地可求得

$$P(1/2)=P\left(\frac{W(\boldsymbol{x})-\nu_2}{\Delta}\geqslant\frac{\Delta}{2}\middle|\boldsymbol{x}\in\pi_2\right)=\Phi\left(-\frac{\Delta}{2}\right)$$

其中 $\Phi(\cdot)$ 为一元标准正态分函数，于是总的误判概率为

$$\text{TPM}=\frac{1}{2}\left[\Phi\left(-\frac{\Delta}{2}\right)+\Phi\left(-\frac{\Delta}{2}\right)\right]=\Phi\left(-\frac{\Delta}{2}\right)\ \blacksquare$$

在一般情况下，总体参数是未知的，它们通过样本来估计，此时线性判别函数便是(2.22)式，即 Anderson 线性判别函数

$$W(\boldsymbol{x})=\left(\boldsymbol{x}-\frac{1}{2}(\bar{\boldsymbol{x}}^{(1)}+\bar{\boldsymbol{x}}^{(2)})\right)'S^{-1}(\bar{\boldsymbol{x}}^{(1)}-\bar{\boldsymbol{x}}^{(2)})$$

它的分布就不是一元正态了，但可以证明（参看 [7]）当 $\boldsymbol{x}\in\pi_1$，$n_1\to\infty$ 时 $W(\boldsymbol{x})$ 的极限分布为 $N\left(\frac{\Delta^2}{2},\ \Delta^2\right)$。当 $\boldsymbol{x}\in\pi_2, n_2\to\infty$ 时 $W(\boldsymbol{x})$ 的极限分布为 $N\left(-\frac{\Delta^2}{2},\ \Delta^2\right)$。其中

$$\Delta^2=(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}).$$

如果总体不是正态情况，有时变得十分复杂。因此在实用上我们引进一个所谓的**明显误判率** (apparent error rate)：将两个

总体的样本代入判别函数进行回报，判错的样本个数占总样本的比例。

设总体 π_1 和 π_2 各有容量为 n_1 和 n_2 的样本，记

n_{1m} = π_1 中的样本判入 π_2 的个数

n_{2m} = π_2 中的样本判入 π_1 的个数

则明显误判率为

$$\text{APER} = \frac{n_{1m} + n_{2m}}{n_1 + n_2} \tag{2.38}$$

明显误判率直观且容易计算。但它作为真正的总误判概率的一个估计，总是偏低的，这是因为判别函数的建立正是利用了这些样本的信息，反过来又对这些样本进行判别，其效果当然会比判别其它个体好。要解决明显误判率偏低的问题除非样本容量 n_1, n_2 都很大，否则这个问题无法从根本上解决。现在已提出一种刀切法或者称为 Lachenbruch 的保留法来估计真实的误判概率。具体做法是：

1° 从总体 π_1 的样本开始，剔除一个样品用剩余的 n_1-1 个样品和总体 π_2 的 n_2 个样品建立判别函数；

2° 用 1° 建立的判别函数去判别刚剔除的那个样品；

3° 重复 1° 和 2° 直到 π_1 的样品都被判别过，用 n_{1m}^* 记错判的个数；

4° 对 π_2 重复 1°，2°和 3° 的作法，并用 n_{2m}^* 记错判的个数。

我们用 $\dfrac{n_{1m}^* + n_{2m}^*}{n_1 + n_2}$ 来作为真实总误判概率的估计。容易看出刀切法的缺点是计算量太大。但对于各种判别规则，如果编制了专门的程序，那么此法还是可行的。

对于多总体的情况明显误判率为

$$\text{APER} = \sum_{j=1}^{q} n_{jm} \Big/ \sum_{j=1}^{q} n_j \tag{2.39}$$

其中 n_j 表示第 j 个总体的样本容量，n_{jm} 表示第 j 个总体的样本

错判到其它总体的个数。

§7.3 Fisher 判 别

(一) 两总体的情况

首先考虑两总体的 Fisher 的判别问题。设两个 p 维总体为 π_1 和 π_2，且都有二阶矩存在。Fisher 的判别思想是变换多元观察 x 到一元观察 y，使得由总体 π_1 和 π_2 产生的 y 尽可能的分离开来。例如 π_1 和 π_2 为二维总体，如图 3.1(a) 所示"×"为 π_1 的点，"▲"为 π_2 的点，它们分别在 x_1 轴和 x_2 轴上的投影有很大一部分是重叠在一起的，故要用原变量 x_1，x_2 的取值范围来把 π_1 与 π_2 分离开来是困难的。Fisher 提出把 y 取做 $x=(x_1,x_2)'$ 的线性组合，即 $y=c_1x_1+c_2x_2$ 它是三维空间中的一个平面 P，只要适当的选取 c_1 与 c_2 使得 π_1 上的点与 π_2 上的点投影在 P 平面上尽可能的分离开来，即在 y 轴上尽可能分离开来，如图 3.1 的 (b)π_1 的 5 个点与 π_2 的 5 个点在 y 轴上已完全分离开了。

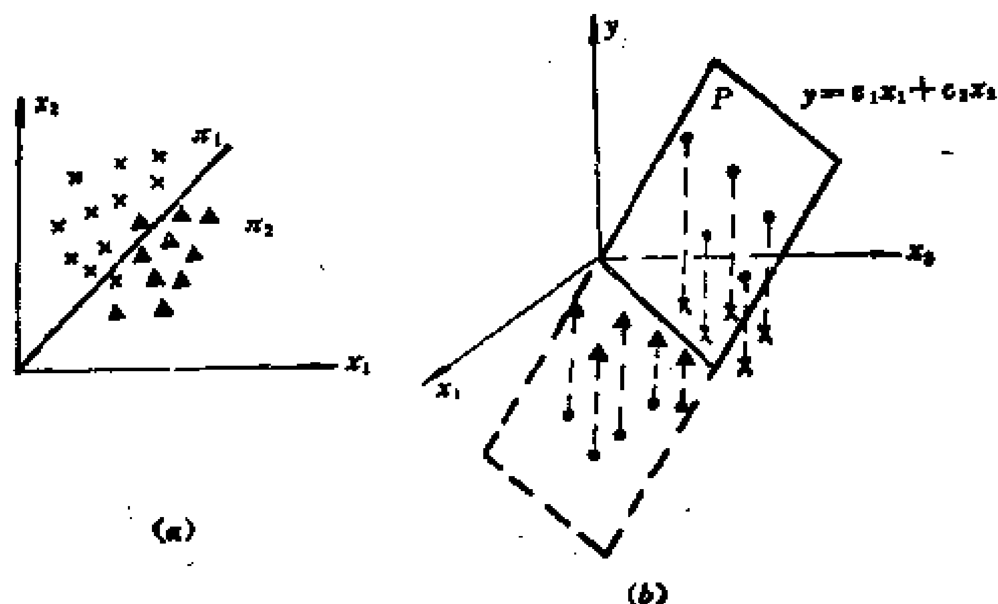


图 3.1

设在 p 维的情况下，x 的线性组合为

$$y=l'x \tag{3.1}$$

其中 $\boldsymbol{l}$ 为 p 维实向量。设 π_1 和 π_2 的均值向量分别为 $\boldsymbol{\mu}^{(1)}$ 和 $\boldsymbol{\mu}^{(2)}$，且有公共的协方差矩阵 $\Sigma(\Sigma > 0)$。那么线性组合 $y = \boldsymbol{l}'\boldsymbol{x}$ 的均值为

$$\mu_{1y} = E(y/\boldsymbol{x} \in \pi_1) = \boldsymbol{l}'\boldsymbol{\mu}^{(1)} \tag{3.2}$$

$$\mu_{2y} = E(y/\boldsymbol{x} \in \pi_2) = \boldsymbol{l}'\boldsymbol{\mu}^{(2)} \tag{3.3}$$

其方差为

$$\sigma_y^2 = \mathrm{var}\, y = \boldsymbol{l}'\Sigma\boldsymbol{l} \tag{3.4}$$

考虑比

$$\frac{(\mu_{1y} - \mu_{2y})^2}{\sigma_y^2} = \frac{[\boldsymbol{l}'(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})]^2}{\boldsymbol{l}'\Sigma\boldsymbol{l}} = \frac{(\boldsymbol{l}'\boldsymbol{\delta})^2}{\boldsymbol{l}'\Sigma\boldsymbol{l}} \tag{3.5}$$

其中 $\boldsymbol{\delta} = \boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$ 为两总体均值向量差，根据 Fisher 的思想我们要选择 $\boldsymbol{l}$ 使得(3.5)式达到最大。

定理 3.1 $\boldsymbol{x}$ 为 p 维随机向量，设 $y = \boldsymbol{l}'\boldsymbol{x}$，当选取 $\boldsymbol{l} = c\Sigma^{-1}\boldsymbol{\delta} = c\Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$($c \neq 0$ 的常数)时，(3.5)式达到最大。

特别当 $c = 1$ 时，线性函数

$$y = \boldsymbol{l}'\boldsymbol{x} = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}\boldsymbol{x} \tag{3.6}$$

称为 Fisher **线性判别函数**。

证明：由第一章中的定理 3.13 便得证。▌

令

$$K = \frac{1}{2}(\mu_{1y} + \mu_{2y}) = \frac{1}{2}(\boldsymbol{l}'\boldsymbol{\mu}^{(1)} + \boldsymbol{l}'\boldsymbol{\mu}^{(2)})$$

$$= \frac{1}{2}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) \tag{3.7}$$

定理 3.2 利用上面的记号，取 $\boldsymbol{l}' = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}$ 则有

$$\begin{aligned} \mu_{1y} - K > 0 \\ \mu_{2y} - K < 0 \end{aligned} \tag{3.8}$$

证明：当 $\boldsymbol{x} \in \pi_1$ 时，$\mu_{1y} = \boldsymbol{l}'\boldsymbol{\mu}^{(1)} = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}\boldsymbol{\mu}^{(1)}$ 因此有

$$\mu_{1y} - K = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}\boldsymbol{\mu}^{(1)}$$

$$- \frac{1}{2}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$$

$$= \frac{1}{2}[(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})] > 0$$

上式的不等号是利用到 $\Sigma^{-1} > 0$。

同样，当 $\boldsymbol{x} \in \pi_2$ 时，$\mu_{2y} = \boldsymbol{l}'\boldsymbol{\mu}^{(2)} = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}\boldsymbol{\mu}^{(2)}$，因此有

$$\mu_{2y} - K = -\frac{1}{2}[(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) < 0 \quad \blacksquare$$

由定理 3.2 我们得到如下的 Fisher 判别规则：

$$\left.\begin{array}{l} \text{当 } (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}\boldsymbol{x} \geqslant K \text{ 时，判 } \boldsymbol{x} \in \pi_1 \\ \text{当 } (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}\boldsymbol{x} < K \text{ 时，判 } \boldsymbol{x} \in \pi_2 \end{array}\right\} \tag{3.9}$$

在实际中若出现 $(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})\Sigma^{-1}\boldsymbol{x} = K$ 时，则可将 $\boldsymbol{x}$ 判给 π_1，也可判给 π_2，但要慎重处理，为了方便我们把它归给 π_1。注意到(3.9)中 K 是由(3.7)所定义，那么有

$$W(\boldsymbol{x}) = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}\boldsymbol{x} - K$$

$$= (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$$

$$= \left(\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})\right)'\Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) \tag{3.10}$$

这样判别规则(3.9)便可写成

$$\left.\begin{array}{l} \text{当 } \boldsymbol{x} \text{ 使得 } W(\boldsymbol{x}) \geqslant 0 \text{ 时，判 } \boldsymbol{x} \in \pi_1 \\ \text{当 } \boldsymbol{x} \text{ 使得 } W(\boldsymbol{x}) < 0 \text{ 时，判 } \boldsymbol{x} \in \pi_2 \end{array}\right\} \tag{3.11}$$

将(3.10)与(1.9)、(2.20)比较，它们完全是一样的，且当 Bayes 判别的两总体(正态总体)的先验概率相等、误判损失相等时，三种判别除了判别函数 $W(\boldsymbol{x})$ 相同外，判别限都是 0。

当总体的参数未知时，我们仍然同前几节一样用样本来对 $\boldsymbol{\mu}^{(1)}$、$\boldsymbol{\mu}^{(2)}$ 及 Σ 进行估计。注意到这里的 Fisher 判别与最小距离判别一样不需要知道总体的分布类型，但两总体的均值向量必须

有显著的差异才行，否则判别无意义。如果总体为正态，均值向量的差异性检验可按第五章所介绍的方法进行。

(二) 多个总体的情况

现在我们将两个总体的 Fisher 判别推广到多个总体。两个总体的 Fisher 判别，是将 p 个指标 $\boldsymbol{x}=(x_1,\cdots,x_p)'$ 综合成一个指标 $y=\boldsymbol{l}'\boldsymbol{x}$ 来代替 p 个原变量进行判别，而且要求两个总体的点在 y 上尽可能地分离开来。推广到多个总体的情况，我们想用少数几个综合指标 $y_i=\boldsymbol{l}_i'\boldsymbol{x}$，$i=1,2,\cdots,r$。且 $r<p$，来代替原变量进行多总体的判别，而且要求 q 个总体在这 r 个综合指标上尽可能地分离开。

设 q 个总体 π_i，$i=1,2,\cdots,q$。有正定的公共协方差矩阵 Σ，π_i 的均值向量为 $\boldsymbol{\mu}^{(i)}$。记

$$\bar{\boldsymbol{\mu}}=\frac{1}{q}\sum_{i=1}^{q}\boldsymbol{\mu}^{(i)},\quad B=\sum_{i=1}^{q}(\boldsymbol{\mu}^{(i)}-\bar{\boldsymbol{\mu}})(\boldsymbol{\mu}^{(i)}-\bar{\boldsymbol{\mu}})' \tag{3.12}$$

考虑线性组合 $y=\boldsymbol{l}'\boldsymbol{x}$，$\boldsymbol{l}$ 为 p 维实向量，对于总体 π_i 相对应于 y 的均值和方差为

$$\mu_{iy}=E(y/\boldsymbol{x}\in\pi_i)=E(\boldsymbol{l}'\boldsymbol{x}|\boldsymbol{x}\in\pi_i)=\boldsymbol{l}'\boldsymbol{\mu}^{(i)} \tag{3.13}$$

$$\sigma_y^2=\mathrm{var}(y/\boldsymbol{x}\in\pi_i)=\boldsymbol{l}'\Sigma\boldsymbol{l} \tag{3.14}$$

记

$$\bar{\mu}_y=\frac{1}{q}\sum_{i=1}^{q}\mu_{iy}=\boldsymbol{l}'\bar{\boldsymbol{\mu}} \tag{3.15}$$

考虑比

$$\frac{\sum_{i=1}^{q}(\mu_{iy}-\bar{\mu}_y)^2}{\sigma_y^2}=\frac{\sum_{i=1}^{q}(\boldsymbol{l}'\boldsymbol{\mu}^{(i)}-\boldsymbol{l}'\bar{\boldsymbol{\mu}})^2}{\boldsymbol{l}'\Sigma\boldsymbol{l}}=\frac{\boldsymbol{l}'B\boldsymbol{l}}{\boldsymbol{l}'\Sigma\boldsymbol{l}} \tag{3.16}$$

显然(3.16)是(3.5)的推广。我们选择 $\boldsymbol{l}$ 使得(3.16)达到最大，为了方便，我们约定 $\boldsymbol{l}'\Sigma\boldsymbol{l}=1$。

定理 3.3 设 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_s>0$ 为 $\Sigma^{-1}B$ 的 s 个非零特征根，$s\leqslant\min(q-1,p)$（见习题 7.5），$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_s$ 为相应

的特征向量（满足 $e'\Sigma e=1$），则 $l_1=e_1$ 时使得(3.16)达到最大，称 $l_1'x=e_1'x$ 为第一判别函数。除去 $l_1'x$，则 $l_2=e_2$ 是在约束条件 $\mathrm{cov}(l_1'x, l_2'x)=0$ 之下使得 (3.16) 达到最大的解，称 $l_2'x=e_2'x$ 为第二判别函数。类推下去，除去 $l_1'x,\cdots,l_{k-1}'x$，则 $l_k=e_k$ 是在约束条件 $\mathrm{cov}(l_k'x, l_i'x)=0$，$i<k$，$k=2,\cdots,s$ 之下使得 (3.16) 达到最大，称 $l_k'x$ 为第 k 个判别函数。同时 $\mathrm{var}(l_i'x)=1$，$i=1,2,\cdots,s$。

证明　令 $a=\Sigma^{\frac{1}{2}}l$，　　(3.17)

则有

$$a'a=l'\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}l=l'\Sigma l,$$

$$a'\Sigma^{-\frac{1}{2}}B\Sigma^{-\frac{1}{2}}a=l'Bl$$

故求(3.16)的极大值问题化为求

$$\frac{a'\Sigma^{-\frac{1}{2}}B\Sigma^{-\frac{1}{2}}a}{a'a} \tag{3.18}$$

的极大值问题。设 $\lambda_1^*\geqslant\lambda_2^*\geqslant\cdots\geqslant\lambda_s^*>0$ 为 $\Sigma^{-\frac{1}{2}}B\Sigma^{-\frac{1}{2}}$ 的 s 个非零特征根，$e_1^*,e_2^*,\cdots,e_s^*$ 为相应的单位化正交特征向量(即长度为 1)，由第一章定理 3. 14 知 (3.18) 在 $a=e_1^*$ 达到最大值且最大值为 λ_1^*。由(3.17)知它等价于在 $l_1=\Sigma^{-\frac{1}{2}}e_1^*$ 处达到，且有

$$\mathrm{var}(l_1'x)=l_1'\Sigma l_1=e_1^{*\prime}\Sigma^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}e_1^*=e_1^{*\prime}e_1^*=1$$

而 $\max\limits_{a\perp e_1^*}\dfrac{a'\Sigma^{-\frac{1}{2}}B\Sigma^{-\frac{1}{2}}a}{a'a}=\lambda_2^*$，且在 $a=e_2^*$ 处达到，同样由(3.17)知它等价于在 $l_2=\Sigma^{-\frac{1}{2}}e_2^*$ 达到，又

$$\mathrm{cov}(l_2'x,l_1'x)=l_2'\Sigma l_1=e_2^{*\prime}e_1^*=0$$

$$\mathrm{var}(l_2'x)=l_2'\Sigma l_2=e_2^{*\prime}e_2^*=1$$

一般地有

$\max\limits_{\substack{a\perp e_i^*\\ i=1,2,\cdots,k-1}}\dfrac{a'\Sigma^{-\frac{1}{2}}B\Sigma^{-\frac{1}{2}}a}{a'a}=\lambda_k^*$，且在 $a=e_k^*$ 处达到，同样等价于在 $l_k=\Sigma^{-\frac{1}{2}}e_k^*$ 达到，而且对一切 $i<k$，$k=2,\cdots,s$。有

$$\operatorname{cov}(l_k'x, l_i'x) = l_k'\Sigma l_i = e_k^{*\prime} \cdot e_i^* = 0$$
$$\operatorname{var}(l_k'x) = l_k'\Sigma l_k = e_k^{*\prime} e_k^* = 1$$

最后我们还要证明 $l_1, \cdots, l_s$ 为 $\Sigma^{-1}B$ 的特征向量 $e_1, e_2, \cdots, e_s$. 因为 $\Sigma^{-\frac{1}{2}}B\Sigma^{-\frac{1}{2}}e_i^* = \lambda_i^* e_i^*$，两边同时左乘 $\Sigma^{-\frac{1}{2}}$ 得

$$\Sigma^{-1}B(\Sigma^{-\frac{1}{2}}e_i^*) = \lambda_i^*(\Sigma^{-\frac{1}{2}}e_i^*)$$

上式表明了 $\Sigma^{-\frac{1}{2}}e_i^* = l_i$ 为矩阵 $\Sigma^{-1}B$ 的特征向量 $e_i\lambda_i^*$ 是 $\Sigma^{-1}B$ 的特征根 $\lambda_i, i = 1, 2, \cdots, s$. ∎

当总体的均值向量 $\mu^{(i)}$, $i = 1, 2 \cdots, q$. 和 Σ 未知时，与往常一样要用总体的样本来估计它们. 设 $x_1^{(i)}, \cdots, x_{n_i}^{(i)}$ 为总体 π_i 的容量为 n_i 的样本 $(i = 1, 2, \cdots, q)$，样本均值向量和协方差阵分别为

$$\bar{x}^{(i)} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_j^{(i)}, S_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(x_j^{(i)} - \bar{x}^{(i)})(x_j^{(i)} - \bar{x}^{(i)})' \tag{3.19}$$

总的样本均值向量为

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{q} n_i\bar{x}^{(i)} = \frac{1}{n}\sum_{i=1}^{q}\sum_{j=1}^{n_i} x_j^{(i)} \quad 其中 \quad n = \sum_{i=1}^{q} n_i \tag{3.20}$$

而

$$\hat{B} = \sum_{i=1}^{q}(\bar{x}^{(i)} - \bar{x})(x^{(i)} - \bar{x})'$$

$$S = \frac{1}{n_1 + \cdots + n_q - q}\sum_{i=1}^{q}(n_i - 1)S_i = \frac{1}{n - q}A \tag{3.21}$$

即我们用 $\bar{x}^{(i)}$、$\bar{x}$、$\hat{B}$ 和 S 分别来代替 $\mu^{(i)}$、$\bar{\mu}$、B 和 Σ. 此外，注意到(3.21)式中 A 为 S 的一个倍数. 因此. $\max\limits_{l}\dfrac{l'\hat{B}l}{l'Sl}$ 与 $\max\limits_{l}\dfrac{l'\hat{B}l}{l'Al}$ 有相同的解，并且如果 $A^{-1}\hat{B}\hat{e} = \hat{\lambda}\hat{e}$，则 $S^{-1}\hat{B}\hat{e} = (n - q)\hat{\lambda}\hat{e}$，即 $A^{-1}\hat{B}$ 与 $S^{-1}\hat{B}$ 有相同的特征向量 $\hat{e}$，于是可以用 $A^{-1}\hat{B}$ 的特征向量 $\hat{e}_i$ 来表示极值问题的解. 这样我们可以得到多个总体的第 i 个 Fisher 样本判别函数

表 3.1　三个沙岩层的原油样品数据

序号	π_1				
	x_1	x_2	x_3	x_4	x_5
1	3.9	51.0	0.20	7.06	12.19
2	2.7	49.0	0.07	7.14	12.23
3	2.8	36.0	0.30	7.00	11.30
4	3.1	45.0	0.08	7.20	13.01
5	3.5	46.0	0.10	7.81	12.63
6	3.9	43.0	0.07	6.25	10.42
7	2.7	35.0	0.00	5.11	9.00

序号	π_2				
	x_1	x_2	x_3	x_4	x_5
1	5.0	47.0	0.07	7.06	6.10
2	3.4	32.0	0.20	5.82	4.69
3	1.2	12.0	0.00	5.54	3.15
4	8.4	17.0	0.07	6.31	4.55
5	4.2	36.0	0.50	9.25	4.95
6	4.2	35.0	0.50	5.69	2.22
7	3.9	41.0	0.10	5.63	2.94
8	3.9	36.0	0.07	6.19	2.27
9	7.3	32.0	0.30	8.02	12.92
10	4.4	46.0	0.07	7.54	5.76
11	3.0	30.0	0.00	5.12	10.77

序号	π_3										
	x_1	x_2	x_3	x_4	x_5		x_1	x_2	x_3	x_4	x_5
1	6.3	13.0	0.50	4.24	8.27	20	6.7	52.0	0.50	4.80	3.20
2	1.7	5.6	1.00	5.69	4.64	21	9.0	27.0	0.30	3.69	3.30
3	7.3	24.0	0.00	4.34	2.99	22	7.8	29.0	1.50	6.72	5.75
4	7.8	18.0	0.50	3.92	6.09	23	4.5	41.0	0.50	3.33	2.27
5	7.8	25.0	0.70	5.39	6.20	24	6.2	34.0	0.70	7.56	6.93
6	7.8	26.0	1.00	5.02	2.50	25	5.6	20.0	0.50	5.07	6.70
7	9.5	17.0	0.05	3.52	5.71	26	9.0	17.0	0.20	4.39	8.33
8	7.7	14.0	0.30	4.65	8.63	27	8.4	20.0	0.10	3.74	3.77
9	11.0	20.0	0.50	4.27	8.40	28	9.5	19.0	0.50	3.72	7.37
10	8.0	14.0	0.30	4.32	7.87	29	9.0	20.0	0.50	5.97	11.17
11	8.4	18.0	0.20	4.38	7.98	30	6.2	16.0	0.05	4.23	4.18
12	10.0	18.0	0.10	3.06	7.67	31	7.3	20.0	0.50	4.39	3.50
13	7.3	15.0	0.05	3.76	6.84	32	3.6	15.0	0.70	7.00	4.82
14	9.5	22.0	0.30	3.98	5.02	33	6.2	34.0	0.07	4.84	2.37
15	8.4	15.0	0.20	5.02	10.12	34	7.3	22.0	0.00	4.13	2.70
16	8.4	17.0	0.20	4.42	8.25	35	4.1	29.0	0.70	5.78	7.76
17	9.5	25.0	0.50	4.44	5.95	36	5.4	29.0	0.20	4.64	2.65
18	7.2	22.0	1.00	4.70	3.49	37	5.0	34.0	0.70	4.21	6.50
19	4.0	12.0	0.50	5.71	6.32	38	6.2	27.0	0.30	3.97	2.97

$$y_i = \hat{e}_i' x, i = 1,2,\cdots,s. \quad (3.22)$$

其中 $\hat{e}_i$ 为 $\hat{A}^{-1}\hat{B}$ 的第 i 个大的特征根所对应的特征向量，且满足 $\hat{e}_i' S \hat{e}_i = 1$，$\hat{e}_i' S \hat{e}_k = 0$，$(i \neq k)$。

至此我们已讨论了如何建立多个总体的 Fisher 判别函数，其判别规则如何确定呢？在介绍判别规则之前先看一个例子。

例 3.1 从三个不同的地区采集了 56 个原油样品，每个样品测试了 5 个指标：x_1=V（钒），$x_2=Fe^{\frac{1}{2}}$（铁$^{\frac{1}{2}}$），$x_3=Pi^{\frac{1}{2}}$（铍$^{\frac{1}{2}}$），x_4 =（饱和烃）$^{-1}$，x_5 = 芳烃。根据其化学成份这 56 个样品归属于三个沙岩层(三个总体)：

π_1: Wilhelm 沙岩 （$n_1 = 7$ 个样品）

π_2: 低 Mulinia 沙岩 （$n_2 = 11$ 个样品）

π_3: 上沙岩 （$n_3 = 38$ 个样品）

测量的原始数据列于表 3.1。

从表 3.1 的数据计算得

$$\bar{x}^{(1)} = (3.229,\ 6.589,\ 0.303,\ 0.150,\ 11.54)'$$

$$\bar{x}^{(2)} = (4.445\ \ 5.667\ \ 0.344\ \ 0.157\ \ 5.454)'$$

$$\bar{x}^{(3)} = (7.226\ \ 4.634\ \ 0.598\ \ 0.223\ \ 5.768)'$$

$$\bar{x} = (6.180\ \ 5.081\ \ 0.511\ \ 0.201\ \ 6.434)'$$

$$\hat{A} = \begin{pmatrix} 187.575 & & & & \\ 1.957 & 41.789 & & & \\ -4.031 & 2.128 & 3.580 & & \\ 1.092 & -0.143 & -0.284 & 0.077 & \\ 79.672 & -28.243 & 2.559 & -0.996 & 338.023 \end{pmatrix}$$

这里 $s = \min(q-1,p) = \min(2,5) = 2$，$\hat{A}^{-1}\hat{B}$ 有两个正的特征根 4.354 和 0.559。Fisher 的两个线性判别函数为

$$y_1 = 0.312x_1 - 0.710x_2 + 2.764x_3 + 11.809x_4 - 0.235x_5$$

$$y_2 = 0.169x_1 - 0.245x_2 - 2.046x_3 - 24.453x_4 - 0.378x_5$$

以 (y_1,y_2) 组成一个二维的判别空间。将三个总体的三组样本数据和它们的样本均值向量回代到判别函数 y_1，y_2，并且把它们点

在二维判别空间中，由图 3.1 可看出三类分离得比较好。对于一个新样品，自然地想到应将它归给距三类的均值点最近的那一类，由此我们可以来建立多总体的 Fisher 判别规则。

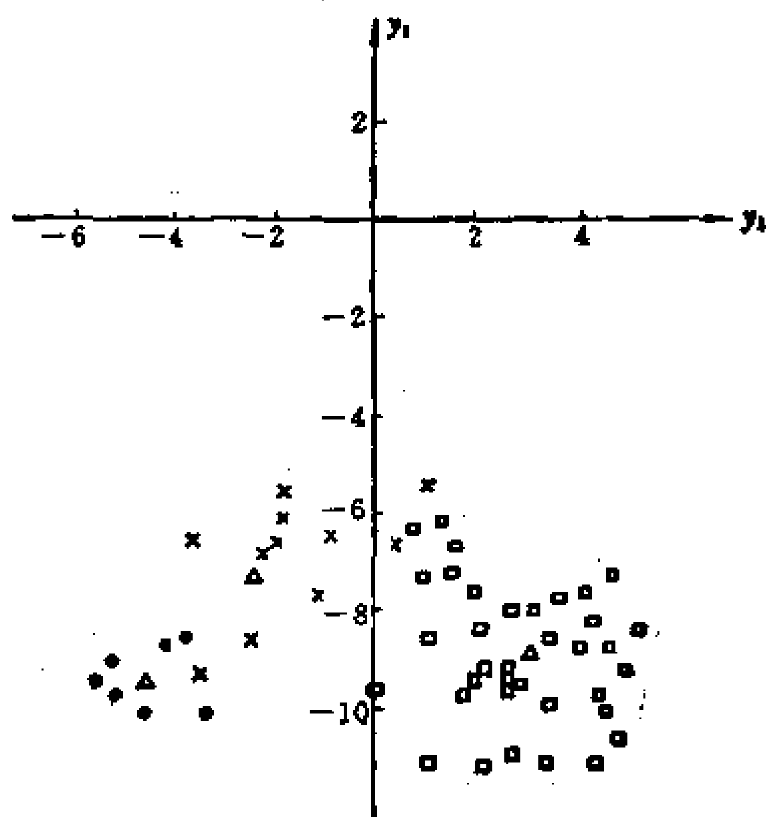


图3.1 ● 表示 π_1；× 表示 π_2；□ 表示 π_3；△ 表示均值

我们已建立的 s 个判别函数为 $y_k = l_k'x$，$k=1,\cdots,s$。它们组成一个判别函数向量 $y=(y_1,\cdots,y_s)'$，其均值向量为 $Ey=(Ey_1,\cdots,Ey_s)'$，对于总体 π_i 而言，有

$$\mu_y^{(i)} = E(y/x\in\pi_i) = (\mu_{iy_1},\cdots,\mu_{iy_s})'$$
$$= (l_1'\mu^{(i)},\cdots,l_s'\mu^{(i)})'$$

对于任给的新样品 x_0，由判别函数它对应着一个 $y_0=(y_{01},\cdots,y_{0s})'$，它到 $\mu_y^{(i)}$ 的欧氏距离平方(因为 $y_1,\cdots,y_s$ 不相关所以采用欧氏距离)为

$$D^2(y_0,\mu_y^{(i)}) = (y_0-\mu_y^{(i)})'(y_0-\mu_y^{(i)}) = \sum_{k=1}^{s}(y_{0k}-\mu_{iy_k})^2 \quad (3.23)$$
$$i=1,2,\cdots,q$$

因此多总体的 Fisher 判别规则为：如果对样品 $\boldsymbol{x}_0$ 有

$$D^2(\boldsymbol{y}_0,\mu_y^{(j)})=\min_{1\leqslant i\leqslant q}D^2(\boldsymbol{y}_0,\mu_y^{(i)}) \tag{3.24}$$

则判 $\boldsymbol{x}_0\in\pi_j$，若满足(3.24)式的标号有 $j_1,j_2,\cdots,j_r$ $(1\leqslant r\leqslant q)$ 则可判 $\boldsymbol{x}_0$ 给 $\pi_{j_1},\cdots,\pi_{j_r}$ 中的任一个。

在 Bayes 判别中，对于有相同协方差矩阵 Σ 的多个正态总体的判别，我们定义了线性判别得分(2.34)，即

$$d_k^*(\boldsymbol{x})=\boldsymbol{x}'\Sigma^{-1}\boldsymbol{\mu}^{(k)}-\frac{1}{2}\boldsymbol{\mu}^{(k)'}\Sigma^{-1}\boldsymbol{\mu}^{(k)}+\ln p_k$$

或等价地为

$$d_k^*(\boldsymbol{x})-\frac{1}{2}\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x}=-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(k)})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(k)})+\ln p_k$$

读者可以证明如下定理。

定理 3.4 设 $y_k=\boldsymbol{l}_k'\boldsymbol{x}$，其中 $\boldsymbol{l}_k=\Sigma^{-\frac{1}{2}}\boldsymbol{e}_k^*$，$\boldsymbol{e}_k^*$ 为 $\Sigma^{-\frac{1}{2}}B\Sigma^{-\frac{1}{2}}$ 的第 k 个大的特征根 λ_k^* 对应的单位化正交特征向量，则有

i) $$\sum_{k=1}^{p}(y_k-\mu_{iy_k})^2=\sum_{k=1}^{p}[\boldsymbol{l}_k'(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})]^2$$
$$=(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})$$
$$=2\left(-d_i^*(\boldsymbol{x})+\frac{1}{2}\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x}+\ln p_i\right)$$

ii) 若 $\lambda_1^*\geqslant\lambda_2^*\geqslant\cdots\geqslant\lambda_s^*>0=\lambda_{s+1}^*=\cdots=\lambda_p^*$，则 $\sum_{k=s+1}^{p}(y_k-\mu_{iy_k})^2$ 与 i 无关，即前 s 个 y_k 或者说 $\sum_{k=1}^{s}(y_k-\mu_{iy_k})^2$ 对判别有贡献。

iii) 对多个具有相同协方差阵的正态总体判别时，若先验概率相等，误判损失也相同，Fisher 判别和 Bayes 判别(2.35)是一样的。

如果前 r 个判别函数 $(r<s)$ 已足够反应 p 个原始变量 $x_1,x_1,\cdots,x_p$ 的变化（参看第九章主成份分析），则只须用前 r 个判

别函数进行判别，这时判别规则为：如果对于样品 x_0 有

$$\sum_{k=1}^{r}(y_{0k}-\mu_{iy_k})^2=\sum_{k=1}^{r}[l'_k(x_0-\mu^{(i)})]^2$$

$$=\min_{1\leqslant j\leqslant q}\sum_{k=1}^{r}[l'_k(x_0-\mu^{(j)})]^2 \qquad (3.25)$$

则判 $x_0\in\pi_i$。若满足 (3.25) 式的标号有 $i_1,\cdots,i_r$ $(1\leqslant r\leqslant q)$ 则可判 x_0 给 $\pi_{i_1},\cdots,\pi_{i_r}$ 中的任何一个。

类似于前面的讨论，若总体的均值向量 $\mu^{(k)}$，$k=1,2,\cdots,q$ 及公共协方差阵 Σ 未知，可用样本均值向量 $\bar{x}^{(k)}$，$k=1,\cdots,q$ 及样本协方差阵 S 来代替。其判别规则与(3.24)、(3.25)一样。

最后，我们从对总体分离的贡献来说明前 r 个判别函数比后面的 $s-r$ 个重要（即第九章中所讨论的前 r 个主成份提供的方差占总方差的大部分）。总体 π_k 的均值向量为 $\mu^{(k)}$，$k=1,2,\cdots,q$，它到总的均值向量 $\bar{\mu}=\frac{1}{q}\sum_{k=1}^{q}\mu^{(k)}$ 的马氏距离为 $(\mu^{(k)}-\bar{\mu})'\cdot\Sigma^{-1}(\mu^{(k)}-\bar{\mu})$，利用 q 个总体的均值向量 $\mu^{(k)}$ 到 $\bar{\mu}$ 的马氏距离之和作为 q 个总体的分离程度的一个度量，即用

$$D^2=\sum_{k=1}^{q}(\mu^{(k)}-\bar{\mu})'\Sigma^{-1}(\mu^{(k)}-\bar{\mu}) \qquad (3.26)$$

来度量 q 个总体的分离程度。

定理 3.5 沿用前面的记号，则 $D^2=\sum_{k=1}^{p}\lambda_k=\sum_{k=1}^{s}\lambda_k$，其中 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_s$ 为 $\Sigma^{-1}B$（或 $\Sigma^{-\frac{1}{2}}B\Sigma^{-\frac{1}{2}}$）的非零特征根，$\lambda_{s+1}=\cdots=\lambda_p=0$，为零特征根。

证明：令 P 为矩阵 $\Sigma^{-\frac{1}{2}}B\Sigma^{-\frac{1}{2}}$ 的特征向量矩阵，它的第 k 列 e_k^* 对应着顺序为第 k 个大的特征根 λ_k。由定理 3.3 知道 $l_k=\Sigma^{-\frac{1}{2}}e_k^*$，$y_k=l'_kx$，考虑

$$y=(y_1,\cdots,y_s,\cdots,y_p)'=P'\Sigma^{-\frac{1}{2}}x$$

而

$$\boldsymbol{\mu}_y^{(k)} = E(\boldsymbol{y}/\boldsymbol{x} \in \pi_k) = P'\Sigma^{-\frac{1}{2}}\boldsymbol{\mu}^{(k)},$$

且

$$\bar{\boldsymbol{\mu}}_y = \frac{1}{q}\sum_{k=1}^{q}\boldsymbol{\mu}_y^{(k)} = P'\Sigma^{-\frac{1}{2}}\bar{\boldsymbol{\mu}}$$

又

$$\begin{aligned}(\boldsymbol{\mu}_y^{(k)} - \bar{\boldsymbol{\mu}}_y)'(\boldsymbol{\mu}_y^{(k)} - \bar{\boldsymbol{\mu}}_y) &= (\boldsymbol{\mu}^{(k)} - \bar{\boldsymbol{\mu}})'\Sigma^{-\frac{1}{2}}PP'\Sigma^{-\frac{1}{2}}(\boldsymbol{\mu}^{(k)} - \bar{\boldsymbol{\mu}}) \\ &= (\boldsymbol{\mu}^{(k)} - \bar{\boldsymbol{\mu}})'\Sigma^{-1}(\boldsymbol{\mu}^{(k)} - \bar{\boldsymbol{\mu}})\end{aligned}$$

从而

$$\begin{aligned}D^2 &= \sum_{k=1}^{q}(\boldsymbol{\mu}^{(k)} - \bar{\boldsymbol{\mu}})'\Sigma^{-1}(\boldsymbol{\mu}^{(k)} - \bar{\boldsymbol{\mu}}) \\ &= \sum_{k=1}^{q}(\boldsymbol{\mu}_y^{(k)} - \bar{\boldsymbol{\mu}}_y)'(\boldsymbol{\mu}_y^{(k)} - \bar{\boldsymbol{\mu}}_y) \\ &= \sum_{k=1}^{q}[(\mu_{ky_1} - \bar{\mu}_{y_1})^2 + (\mu_{ky_2} - \bar{\mu}_{y_2})^2 + \cdots \\ &\quad + (\mu_{ky_p} - \bar{\mu}_{y_p})^2]\end{aligned}$$

注意到对总体 π_k, $\mu_{ky_1} = E(y_1/\boldsymbol{x} \in \pi_k) = \boldsymbol{e}_1^{*\prime}\Sigma^{-\frac{1}{2}}\boldsymbol{\mu}^{(k)}$, $\bar{\mu}_{y_1} = \boldsymbol{e}_1^{*\prime}\Sigma^{-\frac{1}{2}}\bar{\boldsymbol{\mu}}$，所以上式右边第一项可写为

$$\begin{aligned}\sum_{k=1}^{q}(\mu_{ky_1} - \bar{\mu}_{y_1})^2 &= \sum_{k=1}^{q}\boldsymbol{e}_1^{*\prime}\Sigma^{-\frac{1}{2}}(\boldsymbol{\mu}^{(k)} - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}^{(k)} - \bar{\boldsymbol{\mu}})'\Sigma^{-\frac{1}{2}}\boldsymbol{e}_1^* \\ &= \boldsymbol{e}_1^{*\prime}\Sigma^{-\frac{1}{2}}\left[\sum_{k=1}^{q}(\boldsymbol{\mu}^{(k)} - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}^{(k)} - \bar{\boldsymbol{\mu}})'\right]\Sigma^{-\frac{1}{2}}\boldsymbol{e}_1^* \\ &= \boldsymbol{e}_1^{*\prime}\Sigma^{-\frac{1}{2}}B\Sigma^{-\frac{1}{2}}\boldsymbol{e}_1^* = \lambda_1\end{aligned}$$

上式最后一个等号是利用第一章定理 3.13 类似地有

$$\sum_{k=1}^{q}(\mu_{ky_j} - \bar{\mu}_{y_j})^2 = \lambda_j, \ j = 2, \cdots, p$$

从而得到，当 $\lambda_{s+1} = \lambda_{s+2} = \cdots = \lambda_p = 0$ 时有

$$D^2 = \lambda_1 + \lambda_2 + \cdots + \lambda_s + \cdots + \lambda_p = \lambda_1 + \cdots + \lambda_s \quad \blacksquare$$

由定理 3.5 可看出对于第 j 个判别函数 $y_j = l_j'x = e_j^{*\prime}\Sigma^{-\frac{1}{2}}x$ 的均值 μ_{ky_j} 到总均值 $\bar{\mu}_{y_j}$ 的距离平方和为 $\sum_{k=1}^{q}(\mu_{ky_j} - \bar{\mu}_{y_j})^2 = \lambda_j$，它表明了第 j 个判别函数对马氏距离 $D^2 = \sum_{k=1}^{q}(\mu^{(k)} - \bar{\mu})'\Sigma^{-1}\cdot(\mu^{(k)} - \bar{\mu})$的"贡献"为 λ_j。而 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_s$，因此，第一判别函数对 D^2"贡献"最大，然后依次减少。如果前 $r(r < s)$ 特征根之和相对于后 $s-r$ 个特征根之和大得多，则可把后 $s-r$ 个判别函数忽略去而不致使得 D^2 减少很多.这个思想便是第九章中要介绍的主成份的思想。

在本章结束以前，我们再概略介绍一下常用的判别分析中的变量选择问题，与回归分析的自变量筛选问题相类似，如果 x 的变量个数 p 过分大，并由此建立上述的判别函数，显然工作量十分大，而且把一些事实上对判别不起作用的变量也罗列其中。因此我们希望筛选出对判别有用的 x 的部分变量，并由此建立一个比较简单的判别函数，而又基本上保持它原来的判别效果。由此就产生了下面两个问题：i）根据什么标准决定选入的 r 变量对于判别是否重要，ii）如何引进变量和剔除变量。

i）判别准则可如下决定：设有 m 个正态总体 π_i: $N_p(\mu^{(i)}, \Sigma)$ $(i = 1,\cdots,m)$，我们根据 $x^{(i)}$ 的 p 个分量可以对检验 H_0: $\mu^{(1)} = \cdots = \mu^{(m)}$作出相应的检验统计量(见第六章定理 4.1，4.2)

$$T_{(p)} = |E|/|E + B| \tag{3.26'}$$

如果 H_0 未被拒绝，那么根本没必要建立判别函数。只有当 H_0 被拒绝时，我们才认为这 p 个变量对于区分 K 个总体是有意义的，此时才有必要进一步建立判别函数。正如第六章所指出的在 H_0 为真时，

$$T_{(p)} \sim \Lambda(p, n-m, m-1). \tag{3.27}$$

若已知 x 的前 r 个变量对于判别是有显著作用的(即上面 p 改成 r，而相应 H_0 被拒绝)，那么余下的 $p-r$ 个变量是否还有必

要引进呢？这时就应进行检验扣除了前 r 个变量的影响后 $p-r$ 个变量的条件均值是否相等的检验。即检验

$$H_0': \boldsymbol{\mu}_{2\cdot1}^{(1)} = \boldsymbol{\mu}_{2\cdot1}^{(2)} = \cdots = \boldsymbol{\mu}_{2\cdot1}^{(m)} \quad 其中 \quad \boldsymbol{\mu}_{2\cdot1}^{(i)} = E(\boldsymbol{x}_2^{(i)} | \boldsymbol{x}_1^{(i)})$$

从第二章定理 3.2 知

$$\boldsymbol{x}_2^{(i)} | \boldsymbol{x}_1^{(i)} \sim N_{p-r}(\boldsymbol{\mu}_{2\cdot1}^{(i)}, \boldsymbol{\Sigma}_{2\cdot1})$$

其中 $\boldsymbol{\mu}_{2\cdot1}^{(i)} = \boldsymbol{\mu}_2^{(i)} + \Sigma_{22}^{-1}\Sigma_{21}(\boldsymbol{x}_1^{(i)} - \boldsymbol{\mu}_1^{(i)})$，从而可类似推出似然比检验统计量为

$$T_{(p-r,r)} = |E_{2\cdot1}| / |(E+B)_{2\cdot1}| \tag{3.28}$$

其中 $A_{2\cdot1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$，且当 H_0' 为真时

$$T_{(p-r,r)} \sim \Lambda(p-r, n-r-m, m-1). \tag{3.29}$$

并且此时有关系式

$$T_{(p)} = T_{(r)} \cdot T_{(p-r,r)} \tag{3.30}$$

由于实用上总是各个变量依次引进，从而常取 $p-r=1$，此时由第四章(3.2)式有

$$\frac{1-T_{(1,p-1)}}{T_{(1,p-1)}} \cdot \frac{n-m-p+1}{m-1} \sim F(m-1, n-m-p+1) \tag{3.31}$$

(3.31)在逐次引进变量时要反复使用。

在逐次引进变量过程中，当已选人 r 个变量后考虑第 $r+1$ 变量是否需要引进时只须在(3.31)中取 $p=r$ 即可。

ii）确定了判别准则后，具体的引进（或剔除）变量的方法也与回归分析变量筛选完全类似，有前进法，后退法和逐步判别法等，其中逐步判别方法应用较普及，由于逐步判别早已有成熟的程序，实用时只须调用相应程序，输入原始数据，即可找出对判别有影响的变量并建立判别方程。如同回归分析的变量筛选，我们应注意到，运用不同的方法选取变量时选择的结果有时会不同的，需

要详细了解的读者可参阅文献[12].

习 题 七

7.1 设两个二维总体有公共协方差阵，从二总体分别抽取了容量为 $n_1 = 9$，$n_2 = 8$ 的样本其数据如下：

总体 \ 序号		1	2	3	4	5	6	7	8	9
π_1	x_1	20.0	28.5	24.6	26.5	29.0	36.7	36.0	27.6	25.0
	x_2	14.2	8.4	14.8	15.2	11.9	9.6	18.0	9.5	16.0
π_2	x_1	21.6	23.2	17.8	14.4	11.0	18.9	15.4	16.1	
	x_2	7.5	4.5	9.2	7.2	5.5	6.3	8.0	7.8	

(a) 试用最小距离法建立判别函数；

(b) 设 π_1 的先验概率为 $\frac{2}{3}$，π_2 的先验概率为 $\frac{1}{3}$，误判损失相等，二总体均为正态，求出 Bayes 判别区域；

(c) 试比较上述两种判别的明显误判率.

7.2 某科学工作者要建立两个多元总体 π_1 和 π_2 的判别法则，利用他获得的数据已对总体密度 $f_1(x)$，$f_2(x)$ 做出了估计. 设 $C(2/1) = 50$，$C(1/2) = 100$，另外在所有可能的研究对象中，大约有 20% 属于 π_2.

(a) 试导出 ECM 最小判别规则；

(b) 设有一个新样品 x_0，使得 $f_1(x_0) = 0.3$，$f_2(x_0) = 0.5$，x_0 应判入哪个总体.

7.3 某地区将农村经济类型分为三类：π_1——较富裕类型，π_2——中等类型，π_3——较贫困类型. 每种类型以五个指标为依据：x_1 = 土地生产率 = 农村社会总产值/总土地面积（百元/每亩），x_2 = 劳动生产率 = 农村社会总产值/农村劳动力（百元/每个劳动力），x_3 = 人均收入 = 农村经济纯收入/农业人口（百元/每人），x_4 = 费用水平 = 总费用/总收入，x_5 = 农村工业比重 = 农村工业产值/农村社会总产值. 每种类型分别有容量为 $n_1 = 5$，$n_2 = 8$，$n_3 = 4$ 的样本（每个个体以县为单位），其数据如下：

总体	序号	1	2	3	4	5	6	7	8
π_1（较富裕）	x_1	3.85	3.51	4.12	5.01	3.67			
	x_2	6.75	5.73	4.45	4.68	5.84			
	x_3	4.79	4.01	3.68	3.64	4.27			
	x_4	0.85	0.81	0.89	0.78	0.87			
	x_5	0.59	0.60	0.48	0.53	0.64			
π_2（中等）	x_1	3.61	3.65	4.11	2.98	3.21	2.87	3.35	4.00
	x_2	4.05	3.74	4.13	3.69	3.55	3.78	3.81	4.27
	x_3	2.65	2.86	3.15	2.90	3.13	2.60	2.71	2.97
	x_4	0.90	0.91	0.93	0.89	0.88	0.94	0.95	0.90
	x_5	0.45	0.43	0.41	0.41	0.48	0.39	0.38	0.44
π_3（较贫困）	x_1	3.23	3.03	2.54	2.11				
	x_2	4.08	3.21	3.50	2.98				
	x_3	1.85	2.03	1.51	1.07				
	x_4	0.96	0.94	0.97	0.99				
	x_5	0.38	0.25	0.34	0.21				

(a) 设误判损失相等．π_1、π_2 和 π_3 的先验概率分别为 $p_1=\frac{5}{17}$，$p_2=\frac{8}{17}$ 和 $p_3=\frac{4}{17}$，并认为三总体都是正态分布且具有公共协方差阵，试作 Bayes 判别。

(b) 试作 Fisher 判别。

(c) 比较两种判别的明显误判率。

7.4　考虑三个总体的 Bayes 的判别问题，设先验概率、误判损失以及密

		真实总体		
		π_1	π_2	π_3
分类	π_1	$C(1/1)=0$	$C(1/2)=500$	$C(1/3)=100$
	π_2	$C(2/1)=10$	$C(2/2)=0$	$C(2/3)=50$
	π_3	$C(3/1)=50$	$C(3/2)=200$	$C(3/3)=0$
先验概率		$p_1=0.05$	$p_2=0.60$	$p_3=0.35$
$f_i(x_0)$		$f_1(x_0)=0.01$	$f_2(x_0)=0.85$	$f_3(x_0)=2$

度函数在新的样品点 x_0 处的值如上表

(a) 利用(2.28)式，即 ECM 最小的规则，应把 x_0 判给哪一个总体？

(b) 若所有误判损失相等，此时 x_0 应判入哪个总体？并计算其后验概率。

7.5 设 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s > 0$ 为 $\Sigma^{-1}B$ 的 s 个非零特征根，证明 $s \leq \min(q-1, p)$，其中 B 为(3.12)所定义，Σ 为 q 个 p 维总体 $\pi_i(i=1,2,\cdots,q)$ 的公共协方差矩阵 $(\Sigma>0)$。

7.6 证明定理 3.4。

7.7 设总体 $\pi_i(i=1,2,\cdots,q)$ 的概率密度为

$$f_i(x) = \begin{cases} \dfrac{1}{\beta_i} e^{-\frac{x}{\beta_i}} & 0 < x < +\infty \\ 0 & \text{其它} \end{cases}$$

其中 $0<\beta_1<\beta_2<\cdots<\beta_q$ 是未知参数，且设误判损失相等。

(a) 若先验概率相等，即 $p_i = \dfrac{1}{q}, i=1,2,\cdots,q$。则 Bayes 判别区域为：$R_i = (a_{i-1}, a_i]$。其中

$$a_i = \frac{\beta_i \beta_{i+1}}{\beta_{i+1} - \beta_i} \cdot \log \frac{\beta_{i+1}}{\beta_i}, \quad i = 1,2,\cdots,q-1, a_0 = 0$$

$$a_q = +\infty.$$

(提示：由于 $p_i = \dfrac{1}{q}$，$R_i = \{x: f_i(x) > f_k(x)$ 所有 $k \neq i\}$，再注意到 $\dfrac{\beta_i \beta_k}{\beta_k - \beta_i} \log \dfrac{\beta_k}{\beta_i}$ 当 β_i 固定它是 β_k 的上升函数)

(b) 若先验概率成比例，则 Bayes 判别规则相当于将所有个体都判给 π_q。

(提示：只须证明 $R_q = \{x: x>0\}$)

第八章　聚类分析

上一章我们讨论了判别分析，它考虑的是已知 q 个总体 π_i $(i=1,2,\cdots,q)$，对已知的新个体 x 如何判别它属于哪一个总体？而本章讨论的是已知研究对象的一批个体，如何依据它们的 p 个指标(变量)将其分为若干个类型，这就是聚类分析方法，即"物以类聚"的统计方法．分类学是一门古老的学科，从前人们进行分类主要是依靠经验和专业知识，很少利用数学工具．随着科学技术的发展，特别是近代计算机的发展和普及，数学这个有力的工具逐渐被引进到分类学中，形成了数值分类学，由于数值分类在发展的过程中首先是着重于实用，因此相对而言理论上还不太完善，这有待进一步的研究和解决．

本章讲述的聚类分析是数值分类，基本目标是寻找样品或变量的自然类别．对样品的聚类称为 Q 型聚类，对变量的聚类称为 R 型聚类．在第一节中我们先介绍样品之间和变量之间的关系的度量尺度，在后四节中我们介绍聚类分析中最常用的一些方法．

§8.1　相似性与关联性的度量

数值分类法几乎总是对一组个体的若干变量（有时也称为指标或者属性)进行测试，并建立起"相似性"或"相关性"的度量．而个体的若干变量可以是定量的也可以是定性的，定量的可以用数值表示，它包括了连续变量和离散变量，例如个体中某元素的含量(%)就是连续变量；个体中斑点的数目就是离散变量．而定性变量则包括名称的(如土壤的颜色有红壤、黄壤、黑壤等)，次序的(如种子的大小可分为大、中、小等)，状态的(如阴性，阳性等)．定量的和定性的变量可以互相转化，例如定量变量在它取值范围内分割为若干个区间，每个区间根据实际问题可赋于一个名称，此时变

为定性的，相反一个定性变量，它所处的状态，可赋于适当的编码就可化为定量的。当然，关于变量类型的转化还涉及到许多具体问题，这方面的详细讨论可参看[17]。

（一） 距离与相似系数（Q 型）

设每个个体有 p 个变量，因此可把一个个体视为 p 维空间中的一个点，n 个个体就是 p 维空间中的 n 个点，很自然地想到用它们之间的距离来衡量它们之间的靠近程度。在 § 7.1 最小距离判别中，我们已介绍了距离的定义以及列举了一些常用的距离，即对样品点 $\boldsymbol{x}_i$ 与 $\boldsymbol{x}_j$ 有

（1）明考夫斯基距离：

$$d_{ij}=\left[\sum_{\alpha=1}^{p}|x_{\alpha i}-x_{\alpha j}|^{m}\right]^{\frac{1}{m}}\quad m>0。\tag{1.1}$$

取 $m=1,2,\infty$ 分别可得到如下三种距离。

（2）绝对距离：

$$d_{ij}=\sum_{\alpha=1}^{p}|x_{\alpha i}-x_{\alpha j}|\tag{1.2}$$

（3）欧氏距离：

$$d_{ij}=\left[\sum_{\alpha=1}^{p}(x_{\alpha i}-x_{\alpha j})^{2}\right]^{\frac{1}{2}}\tag{1.3}$$

（4）切比雪夫距离：

$$d_{ij}=\max_{1\leqslant\alpha\leqslant p}|x_{\alpha i}-x_{\alpha j}|\tag{1.4}$$

以上的几种距离在 Q 型聚类中是常用的，但各变量的测量值相差悬殊时，要先标准化再计算距离。如果考虑到个体的 p 个变量的相关性等问题，那么我们可采用方差加权距离或马氏距离。

（5）方差加权距离：

$$d_{ij}=\left(\sum_{\alpha=1}^{p}\frac{(x_{\alpha i}-x_{\alpha j})^{2}}{\sigma_{\alpha}^{2}}\right)^{\frac{1}{2}}\tag{1.5}$$

(6) 马氏距离:

$$d_{ij}=(x_i-x_j)'\Sigma^{-1}(x_i-x_j) \tag{1.6}$$

由于将个体视为 p 维空间中的点，当然也就是 p 维空间中的向量。在实际问题中有这种情况：两个物体的大小不一样，但形状极其相似，这两个物体应归同类型，因此引进相似系数。

(7) 相似系数

$$\cos\theta_{x_i,x_j}=\frac{x_i'x_j}{(\|x_i\|\cdot\|x_j\|)^{\frac{1}{2}}}=\frac{\sum_{\alpha=1}^{p}x_{\alpha i}x_{\alpha j}}{\left(\sum_{\alpha=1}^{p}x_{\alpha i}^2\cdot\sum_{\alpha=1}^{p}x_{\alpha j}^2\right)^{\frac{1}{2}}} \tag{1.7}$$

如果个体的各个变量都取正值，则 $0<\cos\theta_{x_i,x_j}\leqslant 1$。当 $\cos\theta_{x_i,x_j}=1$ 时，说明 x_i 与 x_j 完全相似；$\cos\theta_{x_i,x_j}$ 越接近于 1 表明 x_i 与 x_j 越相似。

(二) 匹配系数

在实际问题中，特别是在农业、生物及医学中，个体的 p 个变量都是二态变量，即变量只有两种状态，如医学中的"阴性"和"阳性"，植物分类学中的"有边毛"和"无边毛"，植被考查中某个植物品种的"存在"和"不存在"等等。对于这种二态变量我们引入编码来刻划，即变量处于某一状态时给予编码 1，而处于另一状态时给予编码 0.例如在植被调查中，对两个样地(两个个体) i 与 k 调查了五个品种(五个二态变量)，某品种存在给予编码 1，不存在给予编码 0，调查结果如表 1.1。从表 1.1 中可看出，样地 i 与 k 对五个变量来说有两个(1—1)匹配，一个(0—0)匹配和两个非匹配，即一个(0—1)，一个(1—0)。关于个体只有二态变量的情况其相似性，可以通过定义各种匹配系数来度量，这种具有二态变量的个体相似性问题比较复杂，要根据专业知识加以具体考虑，一般二态变量可分为对称和不对称，若将 0 与 1 同等对待就是对称的，如在植物分类中"有边毛"和"无边毛"是一个很重要的特性，二者必须同等看

表 1.1　i 与 k 两个样地调查结果

个体 \ 种	1	2	3	4	5
i	1	0	0	1	1
k	1	1	0	1	0

待，因此它们是对称的．若 0 与 1 不同等对待就是不对称的，如在植被考查中某个种的"存在"与"不存在"在许多场合中不能等同对待．因此在具体定义具有二态变量的两个个体的匹配系数时要考虑到对称性的问题，对于不对称的情况有时(1—1)匹配比(0—0)匹配的相似迹象更强，在这种情况下扣除一些(0—0)匹配(甚至忽略)是合理的．

为了定义各种匹配系数我们利用 2 × 2 列联表，设个体 i 与 k 具有 p 个二态变量，通过观察便可得表 1.2 所示的个体 i 与 k 的 2 × 2 列联表

表 1.2　个体 i 与 k 的 2×2 列联表

i \ k	1	0	边和
1	a	b	$a+b$
0	c	d	$c+d$
边和	$a+c$	$b+d$	$p=a+b+c+d$

表 1.2 中的 a 表示(1—1)匹配对数，d 表示(0—0)匹配对数，b 和 c 分别表示(1—0)和(0—1)的非匹配对数．关于匹配系数已有十多种定义，表 1.3 列出了九种常见的匹配系数．

(三)　变量的关联性度量(R 型)

对 p 个变量进行聚类，即 R 型聚类，衡量变量之间的关联程度

表1.3　具有 p 个二态变量的两个体匹配系数

匹配系数	说　　明
$\frac{a+d}{p}$	称为简单匹配系数，对(1—1)和(0—0)等同加权，对称情况．
$\frac{2(a+d)}{2(a+d)+b+c}$	对(1—1)和(0—0)匹配双倍加权，对称情况．
$\frac{a+d}{a+d+2(b+c)}$	对非匹配双倍加权，对称情况．
$\frac{a}{p}$	分子上略去(0—0)匹配，非对称情况．
$\frac{a}{a+b+c}$	分子和分母都略去(0—0)匹配，非对称情况．
$\frac{2a}{2a+b+c}$	分子与分母都略去（0—0）匹配，对(1—1)匹配双倍加权，非对称情况．
$\frac{a}{a+2(b+c)}$	分子与分母都略去(0—0)匹配，对非匹配双倍加权，非对称情况．
$\frac{a}{b+c}$	除去(0—0)匹配，(1—1)匹配与非匹配之比，非对称情况．
$\frac{a}{[(a+b)(a+c)]^{\frac{1}{2}}}$	这个系数是 Ochiai 1951 年提出的，非对称情况．

最常用的有相关系数及变量的夹角余弦．

（1）相关系数

设有 n 个个体，每个个体测量了 p 个变量，第 i 个变量 x_i 与第 k 个变量 x_k 的相关系数为

$$r_{ik}=\frac{\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)(x_{kj}-\bar{x}_k)}{\sqrt{\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2\cdot\sum_{j=1}^{n}(x_{kj}-\bar{x}_k)^2}},i,k=1,2,\cdots,p \tag{1.8}$$

(1.8)式的相关系数是大家熟知的. 在某些聚类中负相关系数值往往由它的绝对值来代替.

对于二态变量来说,(1.8)式可以写成更简洁的形式,事实上,通过观察 n 个个体的两个变量 x_i 与 x_k, 便可得到列联表 1.4.

表 1.4 x_i 与 x_k 的列联表

变量 x_k / 变量 x_i	1	0	边和
1	a	b	$a+b$
0	c	d	$c+d$
边和	$a+c$	$b+d$	$n=a+b+c+d$

其中 a 表示变量 x_i 和 x_k 在 a 个个体上同时取 1. 其它 b、c 和 d 有类似地解释. 这样我们就不难得到二态变量 x_i 与 x_k 的相关系数为(见习题 8.2)

$$r_{ik}=\frac{ad-bc}{[(a+b)(c+d)(a+c)(b+d)]^{\frac{1}{2}}} \tag{1.9}$$

在一元统计中,我们知道列联表的独立性检验应用到二态变量时(参看[1]),检验的统计量为 χ^2 统计量:

$$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}\sim\chi^2(1) \tag{1.10}$$

把(1.10)与(1.9)比较,显然有

$$r_{ik}^2 = \frac{1}{n}\chi^2,$$

即说明了对于固定的 n，两个二态变量的相关性大与不独立的概念是一致的。利用两个二态变量的列联表，可以完全类似地构造如表 1.3 所列举的两个二态变量的各种关联系数。

如果两个变量是多态的，例如 x_i 可以取 r 种状态，x_k 可以取 q 种状态，此时它们的列联表为 $r \times q$ 列联表（如表 1.5 所示）。

表 1.5 x_i 与 x_k 的 $r\times q$ 列联表

变量 x_k ╲ 变量 x_i	1	2	…	q	边和
1	n_{11}	n_{12}	…	n_{1q}	$n_{1\cdot}$
2	n_{21}	n_{22}	…	n_{2q}	$n_{2\cdot}$
⋮	⋮	⋮	…	⋮	⋮
r	n_{r1}	n_{r2}	…	n_{rq}	$n_{r\cdot}$
边和	$n_{\cdot 1}$	$n_{\cdot 2}$	…	$n_{\cdot q}$	n

其中

n_{ik} 表示在 n 个个体中变量 x_i 取第 i 种状态，变量 x_k 取第 k 个状态的个数，$n_{i\cdot}$，$n_{\cdot k}$ 分别表示列联表中的行和与列和。同样由一元统计知道 x_i 与 x_k 的独立性检验统计量为

$$\chi^2 = n\sum_{i=1}^{r}\sum_{k=1}^{q}\frac{\left(n_{ik} - \frac{n_{i\cdot}n_{\cdot k}}{n}\right)^2}{n_{i\cdot}n_{\cdot k}} \sim \chi^2[(r-1)(q-1)] \tag{1.11}$$

建立在(1.11)基础上的关联系数也是有很多，下面列举的三个是最常用的：

$$r_{ik}(1) = \left(\frac{\chi^2}{\chi^2 + n}\right)^{\frac{1}{2}} \tag{1.12}$$

$$r_{ik}(2) = \left(\frac{\chi^2}{n \cdot \max(q-1, r-1)}\right)^{\frac{1}{2}} \tag{1.13}$$

$$r_{ik}(3) = \left(\frac{\chi^2}{n \cdot \min(q-1, r-1)}\right)^{\frac{1}{2}} \tag{1.14}$$

(2) 变量的夹角余弦

每一个变量在 n 个个体上都有一个观察值，因此每一个变量可视为 n 维空间中的一个向量，两个变量之间的关系密切程度也可以用它们的夹角余弦表示，变量 x_i 与 x_k 的夹角余弦为

$$\cos\theta_{i,k} = \frac{\sum_{j=1}^{n} x_{i_j} \cdot x_{k_j}}{\sqrt{\left(\sum_{j=1}^{n} x_{i_j}^2\right) \cdot \left(\sum_{j=1}^{n} x_{k_j}^2\right)}} \tag{1.15}$$

根据上述的思想，当然可以计算变量 x_i 与 x_k 在 n 维空间中的距离来衡量它们关系的密切程度，但在实际中很少用.

§8.2 系统聚类法和分解法

本节介绍两种聚类方法，一种是将类由多变到少的系统聚类法；另一种是将类由少变到多的分解聚类法. 为了叙述方便，我们主要讨论 Q 型聚类，而 R 型聚类只通过例子来说明.

(一) 系统聚类法

系统聚类法是目前应用较为广泛的一种聚类方法，有关它的研究极为丰富，而且有专门的程序. 在古老的植物分类学中分类的单位是：门，纲，目，科，属，种. 种是分类的基本单位，分类单位越小它所包括的植物就越少，它们的共同特征就越多. 系统聚类法就是根据这一思想，一开始就把每个个体自成一类，然后把最相似的个体首先聚为小类，再将已聚合的小类按其相似性再聚合，随

着相似性的减弱，最后把一切子类都聚合到一个大类，从而得到一个按相似性大小(或接近程度的大小)联结起来的谱系图.

在系统聚类法的合并过程中涉及到两个新类的距离问题. 类与类之间的距离有许多定义方法，不同的定义法就产生了系统聚类的不同方法. 用 d_{ij} 表示个体 i 与个体 j 的距离，用 π_p 和 π_q 表示两个新类，它们包含的个体数分别用 n_p 和 n_q 表示，它们的重心分别用 $\bar{x}_p$ 和 $\bar{x}_q$ 表示，π_p 和 π_q 之间的距离用 $D_{p,q}$ 表示. 下面给出几个常用的类与类之间的距离定义.

(1) 最小距离.

$$D_{pq}=\min_{i\in\pi_p,j\in\pi_q} d_{ij} \tag{2.1}$$

即用两类中样品之间的距离的最短者作为两类的距离.

(2) 最大距离.

$$D_{pq}=\max_{i\in\pi_p,j\in\pi_q} d_{ij} \tag{2.2}$$

即用两类中样品之间的距离的最长者作为两类的距离.

(3) 重心距离.

$$D_{p,q}=d(\bar{x}_p,\bar{x}_q) \tag{2.3}$$

即用两类的重心之间的距离作为两类的距离.

(4) 类平均距离.

$$D_{p,q}=\frac{1}{n_p n_q}\sum_{i\in\pi_p}\sum_{j\in\pi_q} d_{ij} \tag{2.4}$$

即用两类中所有两两样品之间的距离平均作为两类的距离.

(5) 离差平方和距离.

$$D_{pq}^2=\frac{n_p\cdot n_q}{n_p+n_q}(\bar{x}_p-\bar{x}_q)'(\bar{x}_p-\bar{x}_q) \tag{2.5}$$

离差平方和的方法是由 Ward 首先提出的. 他的思想是来自于方差分析，设 π_p 和 π_q 的离差平方和分别为

$$S_p=\sum_{i\in\pi_p}(x_i-\bar{x}_p)'(x_i-\bar{x}_p),$$

$$S_q = \sum_{x_i \in \pi_q} (x_i - \bar{x}_q)'(x_i - \bar{x}_q),$$

其中 $\bar{x}_p, \bar{x}_q$ 为 π_p 和 π_q 的重心。将 π_p 和 π_q 合并成 π_r，此时 π_r 共有 $n_r = n_p + n_q$ 个个体，π_r 的离差平方和为

$$S_r = \sum_{x_k \in \pi_r} (x_k - \bar{x}_r)'(x_k - \bar{x}_r),$$

其中 $\bar{x}_r$ 为 π_r 的重心。把合并后所增加的离差平方和 $(S_r - S_p - S_q)$ 定义为 π_p 和 π_q 的距离，可以证明(见(二)分解聚类法)增加的离差平方和便是(2.5)所定义的。

有了个体之间的距离和类与类之间距离的定义后，我们便可进行系统聚类，其计算步骤为：

i）n 个个体一开始就作为 n 个类，计算两两之间的距离(或相似系数)构成一个对称阵 $D = \{d_{ij}\}_{n\times n}$，其对角线元素全为 0（相似系数矩阵，对角线元素全为 1）。显然，此时 $D_{pq} = d_{pq}$，将矩阵 D 记为 $D_{(0)}$。

ii）选择 $D_{(0)}$ 中的最小元素（相似系数矩阵则选最大元素），设这个最小元素为 $D_{p,q}$，则将 π_p 和 π_q 合并成一个新类 $\pi_r = \{\pi_p, \pi_q\}$。在 $D_{(0)}$ 中消去 π_p，π_q 所对应的行与列，并加人由新类 $\pi_r = \{\pi_p, \pi_q\}$ 与剩下的其它未聚合的类的距离所组成的一行和一列，得到一个新的距离矩阵 $D_{(1)}$。显然 $D_{(1)}$ 是 $(n-1)\times(n-1)$ 阶矩阵(注：在 $D_{(0)}$ 中如果还有另一个最小元素，则同时可做类似的处理，此时 $D_{(1)}$ 为 $(n-2)\times(n-2)$阶矩阵)。

iii）由 $D_{(1)}$ 出发重复 ii）的做法得到 $D_{(2)}$，再由 $D_{(2)}$ 出发依法炮制，直到 n 个个体聚为一大类为止.

iv）在合并过程中要记下合并时个体的编号以及两类合并时的水平(即两类合并时的距离系数或者相似系数的大小)，并绘成聚类谱系图，根据实际问题确定类的个数。

下面通过例子来说明上述步骤对于类与类之间的距离的不同定义下如何具体实施。

1° 最小距离法。

例 2.1　下面给出了 5 个样品两两之间的欧氏距离矩阵 $D_{(0)}$，试用最小距离法进行系统聚类。

$$
D_{(0)} = \begin{array}{c} \\ \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{array}
\begin{array}{c} \begin{array}{ccccc} \pi_1 & \pi_2 & \pi_3 & \pi_4 & \pi_5 \end{array} \\
\begin{pmatrix} 0 & & & & \\ 4 & 0 & & & \\ 6 & 9 & 0 & & \\ 1 & 7 & 10 & 0 & \\ 6 & 3 & 5 & 8 & 0 \end{pmatrix} \end{array}
$$

因为一开始就将每一个个体作为一类，上面 $D_{(0)}$ 中的 π_i 表示第 i 个个体 $x_i(i=1,2,\cdots,5)$ 且 $D_{i,j}=d_{ij}$。由 $D_{(0)}$ 看出最小的元素为

$$\min\{d_{ij}\} = d_{41} = 1$$

将 $\pi_1 = x_1$，$\pi_4 = x_4$ 合并成新类 $\pi_6 = \{\pi_1, \pi_4\} = \{x_1, x_4\}$，求新类 π_6 与剩下的类的距离

$$D_{6,2} = \min\{d_{1,2}, d_{4,2}\} = \min\{4,7\} = 4$$

$$D_{6,3} = \min\{d_{1,3}, d_{4,3}\} = \min\{6,10\} = 6$$

$$D_{6,5} = \min\{d_{1,5}, d_{4,5}\} = \min\{6,8\} = 6$$

在 $D_{(0)}$ 中消去 π_1 与 π_4 所对应的行与列后，加上 π_6 与 π_2, π_3, π_5 的距离所对应的行与列得 $D_{(1)}$

$$
D_{(1)} = \begin{array}{c} \\ \pi_6 \\ \pi_2 \\ \pi_3 \\ \pi_5 \end{array}
\begin{array}{c} \begin{array}{cccc} \pi_6 & \pi_2 & \pi_3 & \pi_5 \end{array} \\
\begin{pmatrix} 0 & & & \\ 4 & 0 & & \\ 6 & 9 & 0 & \\ 6 & 3 & 5 & 0 \end{pmatrix} \end{array}
$$

$D_{(1)}$ 中的最小元素为 $D_{2,5} = 3$，合并 $\pi_2 = x_2, \pi_5 = x_5$ 得 $\pi_7 = \{\pi_2, \pi_5\} = \{x_2, x_5\}$，计算 π_7 与余下的类 π_3, π_6 的距离

$$D_{7,3} = \min\{D_{2,3}, D_{5,3}\} = \min\{9,5\} = 5$$

$$D_{7,6} = \min\{D_{2,6}, D_{5,6}\} = \min\{4,6\} = 4$$

消去 $D_{(1)}$ 中 π_2，π_5 所在的行与列，加上 π_7 所对应的行与列得 $D_{(2)}$

$$
D_{(2)}=\begin{array}{c}\\ \pi_7\\ \pi_6\\ \pi_3\end{array}\begin{array}{c}\begin{array}{ccc}\pi_7 & \pi_6 & \pi_3\end{array}\\ \begin{pmatrix}0 & & \\ 4 & 0 & \\ 5 & 6 & 0\end{pmatrix}\end{array}
$$

$D_{(2)}$ 中最小元素为 $D_{6,7}=4$，合并 π_6 与 π_7 得 $\pi_8=\{\pi_6,\pi_7\}=\{x_1,x_4,x_2,x_5\}$，最后计算 π_8 与 π_3 的距离

$$
D_{8,3}=\min\{D_{6,3},D_{7,3}\}=\min\{6,5\}=5
$$

并且合并为一个大类 $\pi_9=\{\pi_8,\pi_3\}=\{x_1,\ x_4,\ x_2,\ x_5,\ x_3\}$。至此，对 5 个样品点利用最小距离的系统聚类完毕。归纳上述聚类的过程可列成一个聚类顺序表（表 2.1），并可画出相应的谱系图（图

表 2.1　五个样品点最小距离法归并顺序表

合并次序	合并的类		合并后类中的元素	合并水平(距离)
1	π_1	π_4	$\pi_6=\{x_1,x_4\}$	1
2	π_2	π_5	$\pi_7=\{x_2,x_5\}$	3
3	π_6	π_7	$\pi_8=\{x_1,x_4,x_2,x_5\}$	4
4	π_8	π_3	$\pi_9=\{x_1,x_4,x_2,x_5,x_3\}$	5

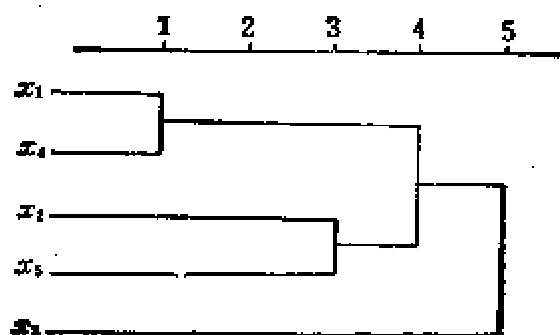


图 2.1　五个样品的最小距离法谱系图

2.1)，图中的横坐标是并类的距离．从图 2.1 可看出在距离为 1 的水平上可分为 4 类：$\{x_1,x_4\}$，$\{x_2\}$，$\{x_5\}$，$\{x_3\}$，在距离为 3 的水平上可分为 3 类：$\{x_1,x_4\}$，$\{x_2,x_5\}$，$\{x_3\}$．在距离为 4 的水平上可分为 2 类：$\{x_1,x_4,x_2,x_5\}$，$\{x_3\}$．最后在距离为 5 的水平上则五个样品全都合并成一类．

最小距离法也可用于 R 型聚类，即对变量进行系统聚类，仍通过例子来说明．

例 2.2　对某地超基性岩的一批样品，测试六个与矿化有关的元素：$x_1=$ 镍，$x_2=$ 钴，$x_3=$ 铜，$x_4=$ 铬，$x_5=$ 硫，$x_6=$ 砷．并计算得它们的相关系数矩阵

$$R_{(0)}=\begin{array}{c}\\ \pi_1\\ \pi_2\\ \pi_3\\ \pi_4\\ \pi_5\\ \pi_6\end{array}\left(\begin{array}{cccccc}\pi_1 & \pi_2 & \pi_3 & \pi_4 & \pi_5 & \pi_6\\ 1 & & & & & \\ 0.8462 & 1 & & & & \\ 0.7579 & 0.9802 & 1 & & & \\ 0.6431 & 0.2419 & 0.1811 & 1 & & \\ 0.5039 & 0.7370 & 0.7210 & -0.3075 & 1 & \\ 0.5603 & 0.4241 & 0.3930 & 0.1998 & 0.6802 & 1\end{array}\right)$$

其中 $\pi_i=x_i$，$i=1,2,\cdots,6$．试用系统聚类的最小距离法对六个变量进行聚类(负相关系数采用绝对值)．

由于采用的是相关系数矩阵，所以应找最大元素并类，在 $R_{(0)}$ 中最大元素为 $R_{3,2}=0.9802$，因此将 π_3 与 π_2 合并为 $\pi_7=\{\pi_3\ \pi_2\}=\{x_3,x_2\}$，计算 π_7 与其它剩下的类的相关系数

$$R_{7,1}=\min\{R_{3,1},R_{2,1}\}=\min\{0.7579,0.8462\}=0.7579,$$

$$R_{7,4}=\min\{R_{3,4},R_{2,4}\}=\min\{0.1811,0.2419\}=0.1811$$

$$R_{7,5}=\min\{R_{3,5},R_{2,5}\}=\min\{0.7210,0.7370\}=0.7210$$

$$R_{7,6}=\min\{R_{3,6},R_{2,6}\}=\min\{0.3930,0.4241\}=0.3930$$

相应地可得到 $R_{(1)}$：

$$
R_{(1)}=\begin{array}{c} \\ x_7 \\ x_1 \\ x_4 \\ x_5 \\ x_6 \end{array}
\begin{array}{c}
\begin{array}{ccccc} x_7 & x_1 & x_4 & x_5 & x_6 \end{array} \\
\left(\begin{array}{ccccc}
1 & & & & \\
0.7579 & 1 & & & \\
0.1811 & 0.6431 & 1 & & \\
0.7210 & 0.5039 & |-0.3075| & 1 & \\
0.3930 & 0.5603 & 0.1998 & 0.6802 & 1
\end{array}\right)
\end{array}
$$

从 $R_{(1)}$ 中找最大的元素 $R_{1,7}=0.7579$，合并 x_1，x_7 得 $x_8=\{x_1,x_7\}=\{x_1,x_2,x_3\}$…如此进行下去，最后将六个变量合并为一个类，详细的合并过程列在表 2.2 中，并且可画出谱系图（图

表 2.2　六个变量的并类顺序表

合并次序	合并的类		合并后的类的元素	合并水平（相关系数）
1	x_2	x_3	$x_7=\{x_2,x_3\}$	0.9802
2	x_1	x_7	$x_8=\{x_2,x_3,x_1\}$	0.7579
3	x_5	x_6	$x_9=\{x_5,x_6\}$	0.6802
4	x_8	x_9	$x_{10}=\{x_2,x_3,x_1,x_5,x_6\}$	0.3930
5	x_{10}	x_4	$x_{11}=\{x_2,x_3,x_1,x_5,x_6,x_4\}$	0.3075

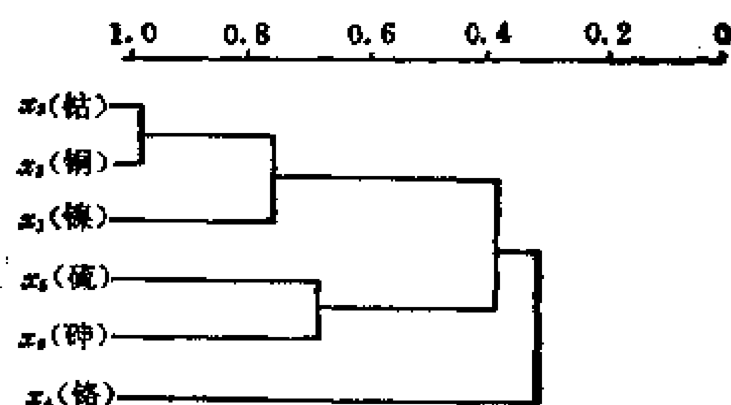


图 2.2　六个元素的谱系图

2.2），横坐标是并类的相关系数．

由上述例子容易看出，计算新类与其它类的距离时，可用如下

的递推公式：

$$D_{r,k}=\min\{D_{p,k},D_{q,k}\} \tag{2.6}$$

其中 $D_{r,k}$ 表示新类 π_r（由类 π_p,π_q 合并的新类）与类 π_k 的距离，$D_{p,k},D_{q,k}$ 分别表示类 π_p,π_q 与类 π_k 的距离。

2° 最大距离法。

最大距离法的计算过程与最小距离法是一样的，只是在计算类与类的距离时采用(2.2)式。

例 2.3 将例 2.1 中的五个样按最大距离法进行系统聚类。

合并的过程在这里就不赘述，只把并类的顺序列于表 2.3 并画出相应谱系图 2.3。

表 2.3 五个样品点最大距离法的合并顺序

合并次序	合并的类		合并后类中的元素	合并水平(距离)
1	π_1	π_4	$\pi_6=\{x_1,x_4\}$	1
2	π_2	π_5	$\pi_7=\{x_2,x_5\}$	3
3	π_6	π_7	$\pi_8=\{x_1,x_4,x_2,x_5\}$	8
4	π_8	π_3	$\pi_9=\{x_1,x_4,x_2,x_5,x_3\}$	10

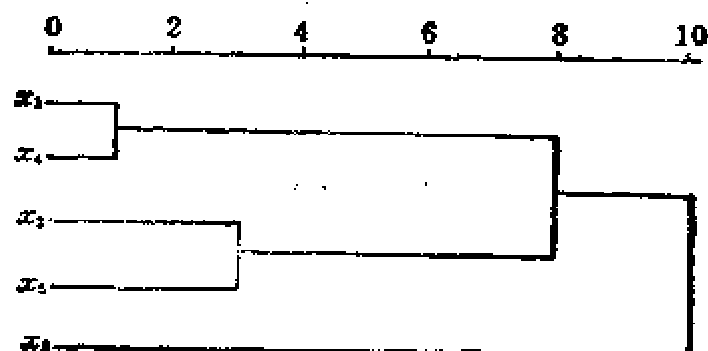


图 2.3 五个样品的最大距离法谱系图

最大距离法同样可进行 R 型聚类。在计算新类 π_r（由类 π_p 和类 π_q 合并的新类）与其它类 π_k 的距离时，同样有递推公式：

$$D_{r,k}=\max\{D_{p,k},D_{q,k}\} \tag{2.7}$$

3° 重心距离法

由于类与类之间的距离是由两类重心的距离来定义，开始时，n 个类之间的距离都是样品之间的距离，随着类的合并新类的重心就要重新计算，因此先导出重心法类间距离的递推公式是方便的。如果某一步类 π_p 与类 π_q 各有 n_p 和 n_q 个样品，其重心分别为 $\bar{x}_p$ 与 $\bar{x}_q$，它们合并成新类 π_r，则 π_r 内有 $n_r = n_p + n_q$ 个样品，其重心为 $\bar{x}_r$，显然

$$\bar{x}_r = \frac{1}{n_r}(n_p\bar{x}_p + n_q\bar{x}_q),$$

采用欧氏距离我们来算 π_r 与其它任意类 π_k 的距离平方，即计算 π_r 的重心 $\bar{x}_r$ 与 π_k 的重心 $\bar{x}_k$ 的距离平方。

$$\begin{aligned}
D_{kr}^2 &= d^2(\bar{x}_k, \bar{x}_r) = (\bar{x}_k - \bar{x}_r)'(\bar{x}_k - \bar{x}_r) \\
&= \left[\bar{x}_k - \frac{1}{n_r}(n_p\bar{x}_p + n_q\bar{x}_q)\right]'\left[\bar{x}_k - \frac{1}{n_r}(n_p\bar{x}_p + n_q\bar{x}_q)\right] \\
&= \bar{x}_k'\bar{x}_k - 2\frac{n_p}{n_r}\bar{x}_k'\bar{x}_p - 2\frac{n_q}{n_r}\bar{x}_k'\bar{x}_q \\
&\qquad + \frac{1}{n_r^2}[n_p^2\bar{x}_p'\bar{x}_p + 2n_pn_q\bar{x}_p'\bar{x}_q + n_q^2\bar{x}_q'\bar{x}_q]
\end{aligned}$$

利用 $\bar{x}_k'\bar{x}_k = \frac{1}{n_r}(n_p\bar{x}_k'\bar{x}_k + n_q\bar{x}_k'\bar{x}_k)$ 及

$$\frac{n_q^2}{n_r^2} = \frac{n_rn_q - n_pn_q}{n_r^2}, \quad \frac{n_p^2}{n_r^2} = \frac{n_rn_p - n_qn_p}{n_r^2}.$$

则有

$$\begin{aligned}
D_{kr}^2 &= \frac{n_p}{n_r}(\bar{x}_k'\bar{x}_k - 2\bar{x}_k'\bar{x}_p + \bar{x}_p'\bar{x}_p) \\
&\quad + \frac{n_q}{n_r}(\bar{x}_k'\bar{x}_k - 2\bar{x}_k'\bar{x}_q + \bar{x}_q'\bar{x}_q) \\
&\quad - \frac{n_pn_q}{n_r^2}(\bar{x}_p'\bar{x}_p - 2\bar{x}_p'\bar{x}_q + \bar{x}_q'\bar{x}_q)
\end{aligned}$$

$$= \frac{n_p}{n_r} D^2_{k,p} + \frac{n_q}{n_r} D^2_{k,q} - \frac{n_p n_q}{n_r^2} D^2_{p,q} \qquad (2.8)$$

(2.8)式就是采用欧氏距离计算的类间重心距离的递推公式。

例 2.4　将例 2.1 中的 5 个样品用重心法进行聚类。

将例 2.1 中的距离系数矩阵 $D_{(0)}$ 中的元素平方并记为 $D^2_{(0)}$:

$$D^2_{(0)} = \begin{array}{c} \\ \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{array} \begin{array}{c} \begin{array}{ccccc} \pi_1 & \pi_2 & \pi_3 & \pi_4 & \pi_5 \end{array} \\ \begin{pmatrix} 0 & & & & \\ 16 & 0 & & & \\ 36 & 81 & 0 & & \\ 1 & 49 & 100 & 0 & \\ 36 & 9 & 25 & 64 & 0 \end{pmatrix} \end{array}$$

在 $D^2_{(0)}$ 中元素 $D_{1,4}=1$ 最小,即 π_1,π_4 最相似,合并它们为 $\pi_6=\{\pi_1,\pi_4\}=\{x_1,x_4\}$, 计算 π_6 与 π_2、π_3、π_5 的类重心距离，此时只须按递推公式(2.8)计算,这里 $n_1=n_4=1, n_6=2$。

$$D^2_{2,6} = \frac{1}{2} D^2_{2,1} + \frac{1}{2} D^2_{2,4} - \frac{1}{2^2} D^2_{1,4}$$

$$= \frac{1}{2} \times 16 + \frac{1}{2} \times 49 - \frac{1}{4} \times 1 = 32.25$$

$$D^2_{3,6} = \frac{1}{2} D^2_{3,1} + \frac{1}{2} D^2_{3,4} - \frac{1}{2^2} D^2_{1,4}$$

$$= \frac{1}{2} \times 36 + \frac{1}{2} \times 100 - \frac{1}{4} \times 1 = 67.75$$

$$D^2_{5,6} = \frac{1}{2} D^2_{5,1} + \frac{1}{2} D^2_{5,4} - \frac{1}{2^2} D^2_{1,4}$$

$$= \frac{1}{2} \times 36 + \frac{1}{2} \times 64 - \frac{1}{4} \times 1 = 49.75$$

从而得到 $D^2_{(1)}$:

$$
D_{(1)}^2 = \begin{matrix} \\ \pi_6 \\ \pi_2 \\ \pi_3 \\ \pi_5 \end{matrix}
\begin{matrix} \pi_6 & \pi_2 & \pi_3 & \pi_5 \\ \left(\begin{matrix} 0 & & & \\ 32.25 & 0 & & \\ 67.75 & 81 & 0 & \\ 49.75 & 9 & 25 & 0 \end{matrix} \right) \end{matrix}
$$

从 $D_{(1)}^2$ 出发重复上述的过程直至五个样品归为一个类。其合并过程由表 2.4 所示。

表 2.4 五个样品点重心法的合并顺序

合并次序	合并的类		合并后类中的元素	合并水平（距离平方）
1	π_1	π_4	$\pi_6 = \{x_1, x_4\}$	1
2	π_2	π_5	$\pi_7 = \{x_2, x_5\}$	9
3	π_6	π_7	$\pi_8 = \{x_1, x_4, x_2, x_5\}$	38.75
4	π_8	π_3	$\pi_9 = \{x_1, x_4, x_2, x_5, x_3\}$	49.56

同样还可画出谱系图。

4° 类平均距离法

类平均距离是采用公式 (2.4)，根据它，我们很容易求得类平均距离的递推公式为

$$
D_{kr} = \frac{1}{n_k n_r} \sum_{i \in \pi_k} \sum_{j \in \pi_r} d_{ij} = \frac{1}{n_k n_r} \left[\sum_{i \in \pi_k} \sum_{j \in \pi_p} d_{ij} + \sum_{i \in \pi_k} \sum_{j \in \pi_q} d_{ij} \right] = \frac{n_p}{n_r} D_{kp} + \frac{n_q}{n_r} D_{kq} \tag{2.9}
$$

利用(2.9)我们就可以进行聚类，其过程与前面几种方法一样。

5° 离差平方和法

离差平方和距离是采用公式(2.5)，根据它并用类似于重心法递推公式的推导方法，便可以得到离差平方和距离的递推公式为

$$
D_{kr}^2 = \frac{n_p + n_k}{n_r + n_k} D_{kp}^2 + \frac{n_q + n_k}{n_r + n_k} D_{kq}^2 - \frac{n_k}{n_r + n_k} D_{pq}^2 \tag{2.10}
$$

利用(2.10)，就可进行聚类，其过程也是与前面几种方法一样。

除了上述介绍的五种系统聚类法之外，还有中间距离法、可变法以及可变类平均法(参见[9]、[17]或[22])，但这三种方法实际应用中较少用。1967 年 Lance 和 Williams 给出了合并前与合并后类与类之间的距离统一递推公式：

$$D_{kr}^2 = \alpha_p D_{k,p}^2 + \alpha_q D_{k,q}^2 + \beta D_{p,q}^2 + \gamma|D_{k,p}^2 - D_{k,q}^2| \quad (2.11)$$

其中 α_p,α_q,β 和 γ 对不同的方法有不同的取值，具体的取值情况列于表 2.5。

表 2.5　八种系统聚类法的参数值

方法	α_p	α_q	β	γ
最小距离法	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
最大距离法	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
重心法	n_p/n_r	n_q/n_r	$-n_p n_q/n_r^2$	0
类平均法	n_p/n_r	n_q/n_r	0	0
离差平方和法	$(n_k+n_p)/(n_k+n_r)$	$(n_k+n_q)/(n_k+n_r)$	$-n_k/(n_k+n_r)$	0
中间距离法	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{4}\leqslant\beta\leqslant 0$	0
可变法	$(1-\beta)/2$	$(1-\beta)/2$	<1	0
可变类平均法	$(1-\beta)n_p/n_r$	$(1-\beta)n_q/n_r$	<1	0

由于有统一公式(2.11)，对编制计算机程序提供了很大的方便。

在结束此段之前，我们再举一个 Q 型系统聚类方法的实际应用例子。

例 2.5　在葡萄酒的质量评比中，由品酒专家分别对每种酒的色、香、味、甜度等十项指标进行打分(10 分为满分)，现有 12 种酒，对每种酒取 25 次样，分别进行打分，然后取平均值，得到每种

表 2.6　十二种葡萄酒得分数据表

酒 \ 指标	1	2	3	4	5	6	7	8	9	10
1	4.58	4.96	6.02	5.43	5.39	4.66	4.23	4.22	4.14	4.39
2	4.56	5.13	6.68	7.15	6.45	3.32	3.64	4.22	4.28	3.59
3	6.13	6.29	7.05	6.76	6.63	6.52	6.81	7.21	6.83	7.35
4	6.08	6.73	6.56	5.80	5.48	5.77	6.00	5.96	5.68	5.75
5	6.20	6.75	6.23	6.58	5.49	5.02	6.31	6.28	6.18	5.76
6	7.40	7.27	6.75	6.70	6.82	6.27	6.25	5.04	6.38	5.16
7	8.05	6.84	6.72	6.34	5.90	7.82	7.02	6.29	6.73	6.11
8	7.80	6.59	6.42	6.18	5.78	7.60	6.57	5.94	5.91	5.21
9	6.97	5.31	5.25	4.83	4.21	6.73	5.06	4.73	4.16	4.25
10	6.87	6.86	6.23	5.56	4.96	5.60	5.77	4.49	3.87	3.34
11	7.60	6.60	5.80	5.32	5.33	6.90	6.35	5.51	5.49	5.67
12	6.96	5.61	4.34	4.28	4.15	6.46	5.70	5.31	4.77	4.19

酒的每项指标的平均得分数据(表 2.6 所示)。

按最小距离法合并的顺序列于表 2.7。其谱系图为图 2.4 所示。

从图 2.4 可以看出,如果我们取合并水平为 3.00,那么十二种酒可以分为三类,第一类为 {1,2},第二类为 {3,4,5,7,8,11,6},第三类为 (9,12,10)。如果我们把每种酒的十项指标的得分再作一次平均,可以看到第一类酒的评分低于 5.0 分,第二类酒的评分几乎都在 6.0 分上下,而第三类酒的评分在 5.0～5.3 分之间. 所以第一类酒最差,第三类酒较差,第二类酒是好酒。用这种聚类方法对酒进行评价是合理的. 如果用离差平方和法或者重心法,仍按三类划分,其结果与最小距离法是一致的。

表 2.7 十二种葡萄酒最小距离法的合并顺序

合并次序	合并的类		合并后类中的元素	合并距离
1	π_4	π_5	$\pi_{13}=\{4,5\}$	1.320
2	π_7	π_8	$\pi_{14}=\{7,8\}$	1.453
3	π_9	π_{12}	$\pi_{15}=\{9,12\}$	1.555
4	π_{14}	π_{11}	$\pi_{16}=\{7,8,11\}$	1.574
5	π_{16}	π_{13}	$\pi_{17}=\{7,8,11,4,5\}$	2.191
6	π_{17}	π_6	$\pi_{18}=\{7,8,11,4,5,6\}$	2.229
7	π_{15}	π_{10}	$\pi_{19}=\{9,12,10\}$	2.687
8	π_1	π_2	$\pi_{20}=\{1,2\}$	2.711
9	π_{18}	π_3	$\pi_{21}=\{7,8,11,4,5,6,3\}$	2.922
10	π_{21}	π_{19}	$\pi_{22}=\{7,8,11,4,5,6,3,9,12,10\}$	3.059
11	π_{22}	π_{20}	$\pi_{23}=\{7,8,11,4,5,6,3,9,12,10,1,2\}$	3.665

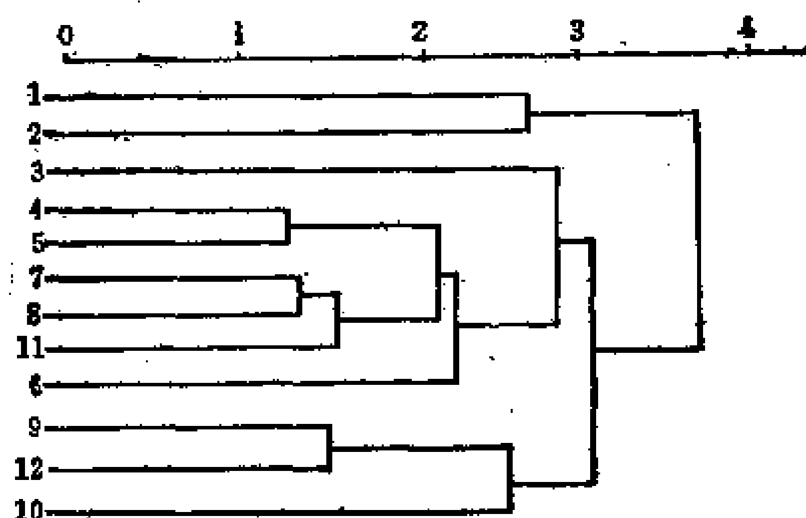


图 2.4 十二种葡萄酒最小距离法谱系图

(二) 分解聚类法

本段介绍将类由少变到多的分解聚类法。一开始是把全部个体当作一类，然后再分为两类，三类，…，直至所有的个体自成一类为止。在这里我们只介绍一种“一分为二”的方法。

设有 n 个样品当作一类记为 π，将它分为两个子类 π_1 和 π_2 且各有 n_1 和 n_2 个样品，记 π,π_1,π_2 的重心为 $\bar{x},\bar{x}_1$ 和 $\bar{x}_2$。"一分为二"的思想方法与上(一)段中 Ward 提出的离差平方和距离的思想是一致的，即方差分析的思想。类 π,π_1,π_2 的离差平方和为

$$S=\sum_{x_i\in\pi}(x_i-\bar{x})'(x_i-\bar{x}) \tag{2.12}$$

$$S_j=\sum_{x_i\in\pi_j}(x_i-\bar{x}_j)'(x_i-\bar{x}_j)\quad j=1,2, \tag{2.13}$$

如果类 π_1 和 π_2 分得合理，应使得 S_1+S_2 尽可能小，或者说使 $S-S_1-S_2$ 尽可能大。容易证明：

$$E=S-S_1-S_2=\frac{n_1\cdot n_2}{n}(\bar{x}_1-\bar{x}_2)'(\bar{x}_1-\bar{x}_2) \tag{2.14}$$

事实上，由于

$$S=\sum_{i=1}^{n}(x_i-\bar{x})'(x_i-\bar{x})=\sum_{i=1}^{n}x_i'x_i-n\bar{x}'\bar{x}$$

$$S_j=\sum_{i=1}^{nj}(x_{ij}-\bar{x}_j)'(x_{ij}-\bar{x}_j)$$

$$=\sum_{i=1}^{nj}x_{ij}'x_{ij}-n_j\bar{x}_j'\bar{x}_j,\quad j=1,2,$$

注意到 $n=n_1+n_2$，及 $\sum_{i=1}^{n}x_i'x_i=\sum_{i=1}^{n_1}x_{i1}'x_{i1}+\sum_{i=1}^{n_2}x_{i2}'x_{i2}$，因此有

$$E=S-S_1-S_2=n_1\bar{x}_1'\bar{x}_1+n_2\bar{x}_2'\bar{x}_2-n\bar{x}'\bar{x} \tag{2.15}$$

将关系式 $n\bar{x}=n_1\bar{x}_1+n_2\bar{x}_2$ 的两边"平方"，得

$$n^2\bar{x}'\bar{x}=n_1^2\bar{x}_1'\bar{x}_1+n_2^2\bar{x}_2'\bar{x}_2+2n_1n_2\bar{x}_1'\bar{x}_2 \tag{2.16}$$

而 $$2\bar{x}_1'\bar{x}_2=\bar{x}_1'\bar{x}_1+\bar{x}_2'\bar{x}_2-(\bar{x}_1-\bar{x}_2)'(\bar{x}_1-\bar{x}_2)$$

将它代入(2.16)得

$$n^2\bar{x}'\bar{x} = n_1(n_1+n_2)\bar{x}_1'\bar{x}_1 + n_2(n_1+n_2)\bar{x}_2'\bar{x}_2 - n_1n_2(\bar{x}_1-\bar{x}_2)'(\bar{x}_1-\bar{x}_2)$$

$$\bar{x}'\bar{x} = \frac{n_1}{n}\bar{x}_1'\bar{x}_1 + \frac{n_2}{n}\bar{x}_2'\bar{x}_2 - \frac{n_1n_2}{n^2}(\bar{x}_1-\bar{x}_2)'(\bar{x}_1-\bar{x}_2)$$

并将它代入(2.15)式便得到(2.14)式。

我们用 $E = S - S_1 - S_2$ 作为目标函数，选择某种分法使 E 达到最大，n 个样品分成二类，一切可能的分法有 $(2^{n-1}-1)$ 种，当 n 稍大时要比较如此多的分类，工作量十分大，于是只好放弃求精确最优解，而去求局部最优解。具体做法如下：

一开始所有 n 个样品均在 π_1 中，然后找一个样品把它归入 π_2 使 E 达到最大，接着再找第二个样品归入 π_2 使 E 达到最大，如此继续下去(注意：某样品一旦归入 π_2 后，该样品在以后的划分中就不再回到 π_1)。

令 $E(1), E(2), \cdots, E(k)$ 表示 π_2 中有一个样品，二个样品，…，k 个样品的 E 值，那么一定存在 k^* 使得

$$E(k^*) = \max_{1\leq k\leq n-1} E(k).$$

于是将前 k^* 次进入 π_2 的样品归为一类，其余 $n-k^*$ 个样品为另一类；这样就把一个类分为两个类。再将上述的步骤应用于子类直至每个样品都自成一类。

例 2.6　表 2.8 给出了 6 个样品的两个变量的测量数据，试用"一分为二"的方法进行聚类。

表 2.8　六个样品的观测数据

变量＼样品	1	2	3	4	5	6
x_1	4	6	0	1	3	5
x_2	10	9	15	14	9	11

开始将六个样品组成一类 $\pi_1^{(0)} = \{x_1, x_2, \cdots, x_6\}$，$\pi_2^{(0)}$ 为空集，将

将 x_1 从 $\pi_1^{(0)}$ 划入 $\pi_2^{(0)}$ 中,相应的目标函数值为

$$E = \frac{n_1 n_2}{n}(\bar{x}_1 - \bar{x}_2)'(\bar{x}_1 - \bar{x}_2) = \frac{5}{6} \cdot \begin{bmatrix} 3 - 4 \\ 11.6 - 10 \end{bmatrix}' \begin{bmatrix} 3 - 4 \\ 11.6 - 10 \end{bmatrix}$$

$$= \frac{5}{6}(1 + 2.56) = 2.97$$

将 x_2 从 $\pi_1^{(0)}$ 划入 $\pi_2^{(0)}$ 中,相应的目标函数值为

$$E = \frac{5}{6} \cdot \begin{bmatrix} 2.6 - 6 \\ 11.8 - 9 \end{bmatrix}' \begin{bmatrix} 2.6 - 6 \\ 11.8 - 9 \end{bmatrix}$$

$$= \frac{5}{6}(11.56 + 7.84) = 16.17$$

类似地将 x_3, x_4, x_5, x_6 分别划到 $\pi_2^{(0)}$ 中，相应的目标函数值为 28.17,14.17,6.57,4.17.其中最大的 E 值是 28.17,即得到 $E(1) = 28.17$，这样就将 x_3 归入 $\pi_2^{(0)}$。于是得到两类为 $\pi_1^{(1)} = \{x_1, x_2, x_4, x_5, x_6\}, \pi_2^{(1)} = \{x_3\}$，此时将 x_3 固定在 $\pi_2^{(1)}$ 中,再重复上述的步骤,即从 $\pi_1^{(1)}$ 中分别将 x_1, x_2, x_4, x_5, x_6 划入 $\pi_2^{(1)}$ 中,计算相应的目标函数值为 8.17，1.42，51.42，9.67，9.67。显然最大者为 51.42，它是将 x_4 划入 $\pi_2^{(1)}$ 的目标函数值，即此时 $E(2) = 51.42$。于是得到两类：$\pi_1^{(2)} = \{x_1, x_2, x_5, x_6\}, \pi_2^{(2)} = \{x_3, x_4\}$。从 $\pi_1^{(2)}$ 和 $\pi_2^{(2)}$ 出发继续做下去,类似的结果列在表 2.9 中。

表 2.9　对六个样品第一分解过程的数据

k	1	2	3	4	5
划出的样品	x_3	x_4	x_6	x_1	x_5
$E(k)$	28.17	51.42	31.14	21.67	16.17

从表 2.9 中可以看出 $E(k^*) = E(2) = \max\limits_{1 \leq k \leq 5} E(k)$，因此对六个样品第一次分解的结果为：

$$\pi_1^{(1)} = \{x_1, x_2, x_5, x_6\}, \quad \pi_2^{(1)} = \{x_3, x_4\}.$$

$E(2)=51.42$ 看作第一次分解的距离水平，记为 $EE(1)$。

将第一次分解结果的两个类分别进行“一分为二”。$\pi_1=\{x_1,x_2,x_5,x_6\}$“一分为二”的结果是：

$$\pi_1'=\{x_1,x_5\},\quad \pi_2'=\{x_2,x_6\}$$

其相应的 E 值为 4.23。而 $\pi_2=\{x_3,x_4\}$ 一分为二的结果是$\{x_3\}$，$\{x_4\}$，其相应的 E 值为 1。显然 4.23>1，此时在水平 $EE(2)=4.23$ 上应分为三类：

$$\pi_1^{(2)}=\{x_1,x_5\},\quad \pi_2^{(2)}=\{x_2,x_6\},\quad \pi_3^{(2)}=\pi_2=\{x_3,x_4\}$$

上面的三类便是第二次分解的结果。

将第二次分解的三个类又进行一分为二，其结果分别为：

$\pi_1^{(2)}=\{x_1,x_5\}$ 分为 $\{x_1\}$，$\{x_5\}$，其 E 值为 1

$\pi_2^{(2)}=\{x_2,x_6\}$ 分为 $\{x_2\}$，$\{x_6\}$，其 E 值为 2.5

$\pi_3^{(2)}=\{x_3,x_4\}$ 分为 $\{x_3\}$，$\{x_4\}$，其 E 值为 1

显然 2.5 最大，此时在水平 $EE(3)=2.5$ 上应分为四类：

$$\pi_1^{(3)}=\{x_1,x_5\},\quad \pi_2^{(3)}=\{x_2\},\quad \pi_3^{(3)}=\{x_6\},\quad \pi_4^{(3)}=\{x_3,x_4\}.$$

再将 $\pi_1^{(3)}$ 与 $\pi_4^{(3)}$ 分别进行一分为二，且 E 值都为 1，最后在水平 $EE(4)=1$ 上将六个样品全部分解完毕。根据 $EE(1)$，$EE(2)$，$EE(3)$，$EE(4)$ 做为分解时的水平（或者用它们的平方根）便可画出谱系图（如图 2.5 所示）。

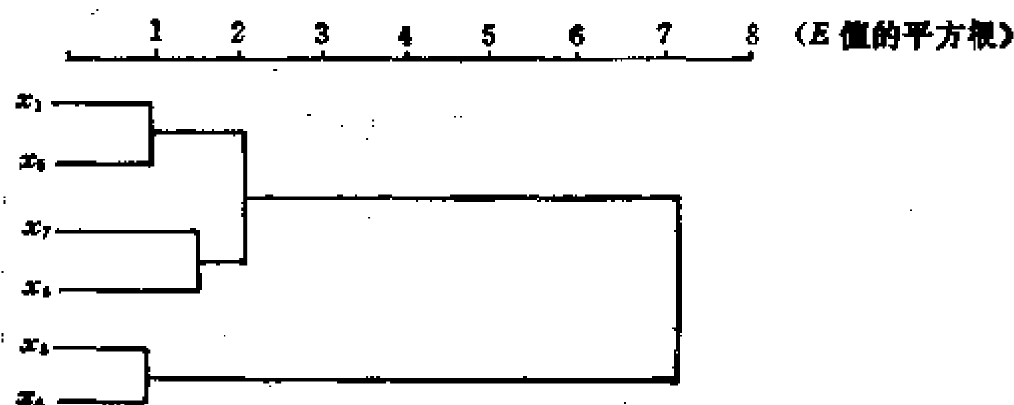


图 2.5　六个样品分解法谱系图

（三） 类的定义及系统聚类的性质

在不同的领域中类的含义是不尽相同的，要给出一个严格的统一定义是不容易的。Rao 1977 年曾给出三种定义法.

定义 2.1 设 h 是给定的正数,对于集合 S 中任二个元素的距离 d_{ij} 有

$$d_{ij} \leqslant h \quad (i,j \in S)$$

称 S 对于阈值 h 组成一个类。

定义 2.2 设 h 是给定的正数,对于集合 S 中元素的距离 d_{ij} 满足

$$\frac{1}{k-1}\sum_{j \in S} d_{ij} \leqslant h, \text{ 对每个 } i \in S$$

其中 k 为 S 中的元素个数。称 S 对于阈值 h 组成一个类。

定义 2.3 设 h 和 $r(>h)$ 是给定两个正数，S 中元素的距离 d_{ij} 满足

$$\frac{1}{k(k-1)}\sum_{i \in S}\sum_{j \in S} d_{ij} \leqslant h, d_{ij} \leqslant r, \ i,j \in S$$

其中 k 为 S 中元素的个数,称 S 对于阈值 h、r 组成一个类.

类似地还可给出如下两个定义。

定义 2.4 给定 $h > 0$，集合 S 中元素的距离 d_{ij} 满足：对于任一个 $i \in S$，存在某个 $j \in S$，使得 $d_{ij} \leqslant h$。称 S 对于阈值 h 组成一个类。

定义 2.5 给定 $h > 0$，将集合 S 任意分成两类 S_1 和 S_2,这两类之间距离 $D(S_1,S_2)$ 满足：$D(S_1, S_2) \leqslant h$。称 S 对于阈值 h 组成一类.

在(一)段中表 (2.5) 给出了类与类之间的八种距离及其递推公式(2.11),类的定义 2.5 中 $D(S_1, S_2)$ 可用八种距离法的任一种(也可用于分解法). 定义 2.1 可用于最大距离法，定义 2.4 可用于最小距离法,定义 2.2 可用于类平均法.

在通常的情况下,不可能确定出一个“最优”的聚类过程,于是

人们提出一种容许聚类过程，令A表示某种在一般情况或者在具有特殊应用范围内的任何合理的聚类过程中均能得到满足的性质。如果一个聚类过程是满足条件A的，我们称该聚类过程（方法)是A容许的(或者称该过程具有性质 A)。1971 年 L. Fisher 和 J. W. Vanness 较详细的讨论了聚类的容许问题，有兴趣的读者可参看[27]。在这里我们只介绍单调性、浓缩与扩张、最小支撑树结构及最优 $W-K$ 分类等性质。

（1）单调性

系统聚类法一开始各样品自成一类，将距离最近的两类合并，这两类距离记为 D_1，第二次合并的最近二类的距离为 $D_2, \cdots$，如果 $D_1 \leqslant D_2 \leqslant \cdots$，则称该系统聚类法具有**单调性**．最小距离法是具有单调性的，事实上，某次把 π_p 和 π_q 类合并为 π_r 时，则 π_r 与其它类 π_k 的距离是

$$D_{r,k} = \min\{D_{p,k}, D_{q,k}\}$$

设下一次并类的距离为 $D_{t,s}$，如果 $t \neq r, s \neq r$，则必有 $D_{t,s} \geqslant D_{p,q}$，否则上一次就应合并 π_t 和 π_s；如果 s, t 有一个等于 r，不妨设 $t = r$，则 $D_{r,s} = \min\{D_{p,s}, D_{q,s}\}$．无论 $D_{r,s}$ 是等于 $D_{p,s}$ $D_{q,s}$ 的哪一个，都有 $D_{r,s} \geqslant D_{p,q}$，所以下次的并类距离 $D_{t,s}$ 不小于上次并类的距离 $D_{p,q}$。这就证明了最小距离法具有单调性．可以证明最大距离法，离差平方和距离法，可变法，可变类平均法以及类平均法也都具有单调性，而重心距离法则没有单调性。

（2）空间的浓缩与扩张

例 2.1 与例 2.3 分别是对同样的五个样品用最小距离法和最大距离法进行聚，它们谱系图分别为图 2.1 和图 2.3，图中的横坐标的范围相差了一倍。而分类水平范围小的方法区别类的灵敏度差，但水平范围太大的方法，灵敏度过高会使没有显著差异变成有显著差异，而归入不同的类。因此灵敏度太低或太高都不好，要适中为好。从这一直观想法引进空间的浓缩与扩张的概念．

设 $A = (a_{ij})$ 与 $B = (b_{ij})$ 是两个元素非负的同阶矩阵，若

$a_{ij} \geqslant b_{ij}$（任意的 i, j），则记 $\mathscr{E}(A) \geqslant \mathscr{E}(B)$（请不要与非负定矩阵 $A \geqslant B$ 混淆），用记号 $\mathscr{E}(A) \geqslant 0$ 表示 A 中的元素非负。

定义 2.6 设 I,II 为两个系统聚类法，在第 i 次合并时相应的距离矩阵记为 $D_i(\mathrm{I})$ 和 $D_i(\mathrm{II})$（$i = 0, 1, \cdots, n-1$），若 $\mathscr{E}(D_i(\mathrm{I})) \geqslant \mathscr{E}(D_i(\mathrm{II}))$，$(i = 0, 1, \cdots, n-1)$，则称 I 比 II 扩张或 II 比 I 浓缩。

我们用 D（小），D（大），D（平），…表示最小距离法，最大距离法，类平均法…合并类时相应的距离矩阵，我们以类平均法为基准可以证明(参看[9])．

$$\mathscr{E}(D\text{（小）}) \leqslant \mathscr{E}(D\text{（平）}) \leqslant \mathscr{E}(D\text{（大）})$$

$$\mathscr{E}(D\text{（重）}) \leqslant \mathscr{E}(D\text{（平）}) \leqslant \mathscr{E}(D\text{（离）}).$$

由此看出类平均法不太扩张也不太浓缩。

*(3) 最小支撑树结构

我们把 n 个样品视为 p 维空间中的 n 个点，点与点之间可用一些边来联结，就构成 p 维空间中的图，因此，自然想到图论中的最小支撑树这个有用的工具．在这里我们着重地讨论最小距离法与最小支撑树的关系。

设 p 维空间有 n 个点，点与点之间定义了距离 d_{ij}，任两点的联线称为**边**，两点间的距离称为**边的长度**。

定义 2.7 两点之间通过一系列的边联系起来，这些边称为**链**，一个链如果是封闭的，则称这个链组成一个**回路**。

定义 2.8 n 个点之间都有链互相联结，则称这些点和链组成了一个**联接图**，没有回路的联接图称为**树**，如果树包含了原联接图的所有点，称这个树为联接图的**支撑树**。

定义 2.9 树的所有边的长度之和叫做树的**重量**，在联接图中具有最小重量的支撑树称为**最小支撑树**，简记为 MST。

设将 n 个点分成两组非空而又不相交的子集 π_p 和 π_q，即 π_p 和 π_q 是 n 个点的一个划分，用(2.1)式定义 π_p 和 π_q 的距离，即

$$D_{p,q}=\min_{i\in\pi_p,j\in\pi_q} d_{ij}$$

令 $C(\pi_p,\pi_q)$ 表示联结 π_p 和 π_q 的一切边的集合，并称它为 π_p, π_q 的截集. 在 $C(\pi_p, \pi_q)$ 中达到极小 $D_{p,q}$ 的边的集合记作 $\lambda(\pi_p,\pi_q)$.

最小支撑树是十分有用的，例如某一城市有 n 个居民点，要建造道路把这 n 个居民点连通，设道路的造价正比于道路长，问应如何修造最合算？它的解正是构造最小支撑树。

为了方便，不妨设 n 个点的任二点的距离不等，事实上若相等，只须其中某边稍微增减一点长度不会影响最小支撑树的计算，在这个假设下 $\lambda(\pi_p,\pi_q)$ 是唯一的.

定理 2.1 设空间 n 个点组成集合 π，则有

i）树的任一边除去后（不包括端点），那么分离成的二部分 π_p, π_q 构成 π 的一个划分，如补上 $C(\pi_p, \pi_q)$ 的任一其它边，构成新的树.

ii）树只有 $n-1$ 条边，且它的任一子集不构成回路，而把任一条不属于树的边加上去，一定形成某一回路.

iii）π 中任意 $n-1$ 条不构成回路的联结图构成一树.

iv）π 中的任一划分 π_p, π_q，则 $\lambda(\pi_p,\pi_q)$ 一定属于 π 的 MST.

v）π 中的任一划分 π_p, π_q，如果对于 π_p 的任一子划分 π_{p_1},π_{p_2} 有 $D_{p_1,p_2}<D_{p,q}$，则 π_p 的 MST 是 π 的 MST 的子集.

证明：i）～ iii）是显然的. 下面证明 iv）和 v）. 设 $\lambda(\pi_p,\pi_q)\bar{\in}$MST，则 MST 一定有 $C(\pi_p,\pi_q)$ 中某一边，根据 i）我们将 $\lambda(\pi_p, \pi_q)$ 代替该边得到新树 T^*，则 T^* 的重量比 MST 轻，得出矛盾这样便证明了 iv). 设某边 $\in\pi_p$ 的 MST，则去掉该边形成 π_p 的一个划分 π_{p_1},π_{p_2}，由 iv)知该边 $=\lambda(\pi_{p_1}, \pi_{p_2})$，又由于 $D_{p,q}=\min(D_{p_1,q},D_{p_2,q})\leqslant D_{p_1,q}$. 由条件 $D_{p_1,p_2}<D_{p,q}$，故

$D_{p_1,p_2} < D_{p_1,q}$. 这说明 π_{p_1} 与 $\pi - \pi_{p_1} = \pi_t$ 的距离 $D_{p_1,t} = \min(D_{p_1,p_2}, D_{p_1,q}) = D_{p_1,p_2}$. 即 $\lambda(\pi_{p_1}, \pi_{p_2}) = \lambda(\pi_{p_1}, \pi_t)$. 由 iv) 知后者属于$\pi$的 MST，即 π_p 的 MST 是π的 MST 的子集。这样便证明了 v).

Kruskal 在 1956 年提出了下列求最小支撑树的算法:

设连结π中任二点的边的全体记为 $C(\pi)$. 显然共有 $n(n-1)/2$ 条.

i) 选取 $C(\pi)$ 中具有最小长度的边 l_1，记 $S_1 = \{l_1\}$,

ii) 在 $C(\pi) - S_1$ 中选取最小长度的边 l_2，记 $S_2 = \{l_1, l_2\}$,

iii) 设已选好了 $S_i = \{l_1, l_2, \cdots, l_i\}$，从 $C(\pi) - S_i$ 中选取 l_{i+1} 使其满足 $S_i \cup l_{i+1}$ 不含回路，同时 l_{i+1} 是 $C(\pi) - S_i$ 中最小长度的边.

iv) 重复 iii) 的作法直到选出 l_{n-1}，从而 $S_{n-1} = \{l_1, l_2, \cdots, l_{n-1}\}$ 构成π的 MST.

定理 2.2 当π中的边长两两不等时按 Kruskal 算法求出的树是π的唯一 MST.

此定理的证明见[28].

下面我们讨论最小距离系统聚类法具有最小支撑树结构.

定理 2.3 设n个样品视为p维空间中的n个点，且它们两两之间的距离互不相等，则在最小距离法中第 $i(1 \leqslant i \leqslant n-1)$ 步合并的两类正是 Kruskal 法中第i步取出的 l_i 所连结的两类，并且此时所有分类 $\pi_1^{(i)}, \cdots, \pi_{n-i}^{(i)}$ 各自的 MST 是π的 MST 的子集.

证明: 最小距离法聚类的第一步是将n个点距离最短的二点归为一类正是 Kruskal 法中取出最小长度的边 l_1. 而类 π_p, π_q 的距离正是 $D_{p,q} = \min\limits_{i \in \pi_p, j \in \pi_q} (d_{ij})$，把 π_p, π_q 合并时，实质上相应于取出边 $\lambda(\pi_p, \pi_q)$. 从而每一次把二类合并的聚类过程可以看成从 $C(\pi)$ 中选出一条边，经过 $n-1$ 次聚类并成一类则相当

于构成了 π 中的某一支撑树. 用归纳法. 设按最小距离法前第 i 次并类相当于 Kruskal 法中取出了 l_i, 则第 $i+1$ 次并类时是从余下的 $n-i$ 类, 记为 $\pi_1^{(i)},\pi_2^{(i)},\cdots,\pi_{n-i}^{(i)}$ 中取出距离最小的二类合并,或者说从所有 $\lambda(\pi_k^{(i)},\pi_s^{(i)})$ ($k\neq s,k,s=1,2,\cdots,n-i$) 中选取最短者 $\bar{l}$, 由最小距离法的单调性知 $\bar{l}>l_i>\cdots>l_1$. 而 $\bar{l}$ 正是 Kruskal 法中的 l_{i+1}. 从而按最小距离法聚类正是求 π 中 MST 的过程. 下面仍利用归纳法来证明第二个结论.当 $i=1$ 时,仅有 l_1 而且是 π 中最短的边, 所以是 π 的 MST 的子集. 设第 i 步得到的类 $\pi_1^{(i)},\cdots,\pi_{n-i}^{(i)}$ 的 MST 是 π 的 MST 的子集为真,当 $i+1$ 时, $\pi_j^{(i+1)}(j=1,2\cdots,n-i-1)$中除了由 l_{i+1} 连结的类 $\pi_l^{(i)}$, $\pi_s^{(i)}$ 之外, 其它的类都与 $\pi_j^{(i)}(j\neq l,j\neq s)$ 一样.因此只须证明 $\pi_l^{(i)}\cup\pi_s^{(i)}$ 的 MST 也是 π 的 MST 的子集即可. 由于 $\pi_l^{(i)}\cup\pi_s^{(i)}$ 的任一个划分 P 与 Q 必有 $D_{P,Q}\leqslant l^{i+1}$, 而 $(\pi_l^{(i)}\cup\pi_s^{(i)})$ 与 $(\pi-\pi_l^{(i)}-\pi_s^{(i)})$ 的距离是大于 l_{i+1}, 由定理 2.1 的 v)知 $\pi_l^{(i)}\cup\pi_s^{(i)}$ 的 MST 必为 π 的 MST 的子集 ▌

*(4) 最优 W-K 分类

设有 n 个样品,欲分成 K 类,设 $P_K=\{\pi_1,\cdots,\pi_K\}$ 是任一种分法,如果能定义一个分类的误差函数 $W(P_K)$, 若 $W(\cdot)$ 越小表示分类越合理,因此我们希望找到一个分法 P_K^* 使得

$$W(P_K^*)=\min_{P_K\in\mathscr{T}_K}W(P_K) \tag{2.17}$$

其中 $\mathscr{T}_K$ 是一切 P_K 的集合,则称 P_K^* 是最优 W-K 分类(在误差函数 $W(P_K)$的意义下),误差函数W的定义有多种多样. 这里仅讨论用类的 MST 的最大边长来定义的误差函数.

设将 n 个样品分成 K 类: $\pi_1,\cdots,\pi_K$, 定义 π_i 的直径 $R=R(\pi_i)$ 为 π_i 的 MST 的最大边长,定义分类的误差函数为

$$W(\pi_1,\cdots,\pi_K)=\max_{1\leqslant i\leqslant K}R(\pi_i) \tag{2.18}$$

定理 2.4 在定理 2.3 的假设下,对一切 K $(1\leqslant K\leqslant n-1)$ 用最短距离法得到的 K 个类 $\pi_1^{(n-K)},\cdots,\pi_K^{(n-K)}$ 是在(2.18)的意

义下的唯一 $W-K$ 分类。

证明 用最小距离法，经过 $n-K$ 次分类，就得到 K 个类：$\pi_1^{(n-K)},\cdots,\pi_K^{(n-K)}$. 由定理 2.3 知道它们的 MST 是 π (n 个样品组成)的 MST 的子集。所以有

$$W(\pi_1^{(n-K)},\cdots,\pi_K^{(n-K)})=\max_{1\leqslant i\leqslant K} R(\pi_i^{(n-K)})=l_{n-K}$$

令 $\pi_1,\cdots,\pi_K$ 是任一个 K 分类，只须证明 $W(\pi_1,\cdots,\pi_K)\geqslant l_{n-K}$ 即可。由最小距离法知 $\pi_1^{(n-K)},\cdots,\pi_K^{(n-K)}$ 任二类间距离最小是 l_{n-K+1}，而 $\{\pi_1,\cdots,\pi_K\}$ 与 $\{\pi_1^{(n-K)},\cdots,\pi_K^{(n-K)}\}$ 不一样，则必存在某一个 $\pi_i(1\leqslant i\leqslant K)$ 它至少包含了 $\{\pi_i^{(n-K)}, i=1,2,\cdots,K\}$ 中两个类的某些元素，不妨设 π_i 包含了元素 x 与 y，其中 $x\in\pi_i^{(n-K)}, y\in\pi_j^{(n-K)}(i\neq j)$. 因为 $\pi_i^{(n-K)}$ 与 $\pi_j^{(n-K)}$ 的距离不小于 l_{n-K+1}，故当联结 x 与 y 的一切链中，其最大边不小于 l_{n-K+1}，这说明了 $R(\pi_i)\geqslant l_{n-K+1}>l_{n-K}$，从而 $W(\pi_1,\cdots,\pi_K)>l_{n-K}$. 而且上述的过程也就是证明这个解的唯一性. ▮

从以上的讨论，可看出最小距离法是具有较多的可容许条件(它满足更多的可容许条件可参看 [27]). 因此，最小距离法在实际应用中很被推崇.

§ 8.3 动态聚类法

我们在上节中讨论了合并的系统聚类法和分解法，它们都有一个共同点，就是样品一旦被归到某个类后就不再变了，这就要求分类的方法比较准确。此外当样品个数较大时，相应地计算量过大。为了弥补它的不足，产生了动态聚类法，它是将样品粗略地先分一下类，然后再按照某种原则进行修正，直至分类比较合理为止。动态聚类法的过程大致上可由图 3.1 所示。

图 3.1 框图的前三个部分都有许多种方法，这些方法按框图进行组合就会得到不同的动态聚类法。我们按前三个框格进行讨论动态聚类。

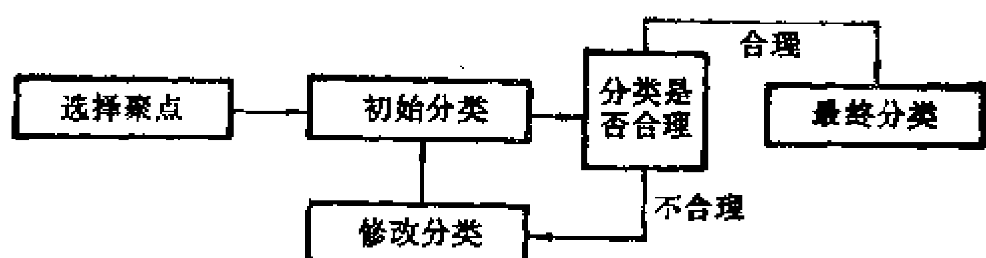


图 3.1 动态聚类的框图表示

(一) 选择聚点的方法

聚点就是一批具有代表性的点，是欲形成类的中心，它的选择决定了初始分类，对最终分类有很大影响，选择时要慎重.选择聚点有如下的办法.

1° 经验选择，如果对研究的对象比较了解，根据以往的经验大体知道样品应如何分类，分成几类(譬如K类)，这样就可以将每一类选择一个有代表性的样品作为聚点.

2° 将n个样品人为地(或随机地)分成K类，以每类的重心作为聚点.

3° 当样品数量n很大时，先从n个样品中随机地抽取一部分，将这部分样品用系统聚类法先进行聚类(聚成K类)，然后计算各类的重心作为聚点.

4° 最小最大原则. 设要将n个样品分成K类，先选择所有样品中相距最远的两个样品点 x_{i_1} 和 x_{i_2} 为前两个聚点. 即

$$d(x_{i_1},x_{i_2})=d_{i_1,i_2}=\max\{d_{i,j}\}$$

然后选择第三个聚点 x_{i_3}，使得 x_{i_3} 与前两个聚点 x_{i_1}，x_{i_2} 的较小距离是所有其余的点与 x_{i_1},x_{i_2} 的较小距离中最大的，然后按同原则选取 x_{i_4}，选取过程可用递推公式表达为:

若已经选了t个聚点 $x_{i_1},\cdots,x_{i_t}$，则$t+1$个聚点选取的原则为

$$\begin{aligned}&\min\{d(x_{i_{t+1}},x_{i_r}),r=1,2,\cdots,t\}\\&=\max\{\min[d(x_j,x_{i_r}):r=1,2,\cdots,t]:j\neq i_1,\cdots,i_t\}\end{aligned}$$

按上述原则选到第K个聚点 x_{i_K} 为止。

以上四种选择聚点的方法是分类数目K是事先固定好的。以下两种方法对K不事先固定。

5° 人为地先选择一个正数 d，把所有样品的重心作为第一个聚点，然后把每个样品输入，如果输入的样品与已确定的聚点的距离都大于 d，则该样品作为一个新的聚点，否则不计。一直把所有样品输完为止。这样便选出了若干个两两距离都大于 d 的聚点。

6° 密度法。人为地定两个正数 d_1 和 $d_2(d_2 > d_1)$，以每个样品点为球心，以 d_1 为半径，落在该球内的样品数（不包括球心的样品）称为该样品的密度。首先选取具有最大密度的样品点为第一聚点，再选择具有次大密度的样品点计算它与第一聚点的距离，如果距离大于 d_2，该点作为第二聚点，否则取消。按此法从密度大到小的顺序选下去，每一次选取的聚点都必须和已选好的聚点的距离大于 d_2。这样便得到了若干个两两距离都大于 d_2 的聚点。

（二） 初始分类的方法

有了聚点后就要进行初始分类，初始分类也有许多方法。当然初始分类也可以和选择聚点同时进行（如下面要介绍的3°），也可以不要聚点直接进行初始分类（如下面要介绍的4°）。

1° 定义样品点之间的距离，每个样品向最靠近的聚点归类。

2° 每个聚点自成一类，样品依次输入，每进一个样品将它归入最近的聚点的类，并计算该类的重心，以这个重心代替原来的聚点，再输入下一个样品，并按同样的原则处理。

3° 先人为地定一个正数 d，选择 $\pi_1=\{x_1\}$，如果 x_2 与 x_1 的距离 $d_{2,1}<d$，则把 x_2 归入 π_1，否则 x_2 自成一类 $\pi_2=\{x_2\}$，当某一步轮到 x_i 输入时，如果已形成K类：$\pi_1, \cdots, \pi_K$，每一类第一次进去的样品记作 $x_{i_1}, \cdots, x_{i_K}$（显然 $i_1=1$），若 $\min\limits_{1\leqslant j\leqslant K} d_{i i_j}>d$，则将 x_i 另立新类 $\pi_{K+1}=\{x_i\}$。

此法的优点是速度快，缺点是分类的结果与样品的次序有关，前面的类容易比后面的类收集到更多的样品。

4° 对样品中的每个变量先人为地定一个数 $d_j(1 \leqslant j \leqslant p)$，令 $x'_{ij} = [x_{ij}/d_j]$，其中 $[x]$ 表示不超过 x 的整数，初始分类的原则是对样品 x_i 与 x_l 归入同一类当且仅当 $x'_{ij} = x'_{lj}$ $(j = 1, 2, \cdots, p)$.

（三） 修改分类的方法

样品进行了初始分类后，对于不合理的分类要进行判断和修正，这也有许多方法，下面我们通过例子来介绍两种方法。

1° 按批修改法

其计算步骤如下：

i）定义样品之间的距离，选择一批聚点（相对应于图 3.1 的第一个框格：选择聚点）。

ii）将样品按最近聚点归类（相对应于图 3.1 的第二个框格：初始分类）。

iii）计算每类的重心作为新的聚点，如果新的聚点与原来的聚点重合，修改过程终止，否则回到 ii）（相对应于图 3.1 的第三个框格：修改分类）

例 3.1 从 12 个不同的地区测定了某树种的平均发芽率（x_1）和发芽势（x_2），其数据如表 3.1，采用欧氏距离，将这 12 个地区的种子用按批修改法聚为两个类。

i）选择聚点，按经验分为两类，并选取 $x_5 = (0.688, 0.605)'$ 和 $x_{12} = (0.777, 0.723)'$ 为聚点。

ii）初始分类，每个样品向最靠近的聚点归类。顺序计算 x_i，$i \neq 5, 12$. 与聚点 x_5, x_{12} 的距离，如 x_1 与 x_5，x_{12} 的欧氏距离为

$$d_{1,5} = [(0.707 - 0.688)^2 + (0.385 - 0.605)^2]^{\frac{1}{2}} = 0.221$$

$$d_{1,12} = [(0.707 - 0.777)^2 + (0.385 - 0.723)^2]^{\frac{1}{2}} = 0.345$$

由于 $d_{1,5} < d_{1,12}$，所以将 x_1 归入 x_5，同样其余的如法炮制，其

表 3.1 十二个地区某树种发芽情况

地区号	1	2	3	4	5	6
x_1	0.707	0.600	0.693	0.717	0.688	0.533
x_2	0.385	0.433	0.505	0.343	0.605	0.380
地区号	7	8	9	10	11	12
x_1	0.877	0.513	0.815	0.633	0.740	0.777
x_2	0.713	0.353	0.675	0.465	0.580	0.723

结果列在表 3.2. 即得到了初始分类为:

$$\pi_1^{(0)} = \{x_1 \sim x_6, x_8, x_{10}, x_{11}\}, \quad \pi_2^{(0)} = \{x_7, x_9, x_{12}\}.$$

iii) 修改初始分类,计算 $\pi_1^{(0)}$ 和 $\pi_2^{(0)}$ 的重心,即两个初始分类的样品点的平均

$$\bar{x}_1^{(0)} = (0.647, 0.450)', \quad \bar{x}_2^{(0)} = (0.823, 0.704)'$$

它们与原来的聚点不同,将重心 $\bar{x}_1^{(0)}, \bar{x}_2^{(0)}$ 作为新的聚点, 将 $x_i(i = 1, 2, \cdots, 12)$ 逐个输入, 按最靠近聚点归类, 其结果列在表 3.3. 即得到第一次按批修改的两个类:

$$\pi_1^{(1)} = \{x_1 \sim x_6, x_8, x_{10}\}, \quad \pi_2^{(1)} = \{x_7, x_9, x_{11}, x_{12}\}$$

计算 $\pi_1^{(1)}$ 与 $\pi_2^{(1)}$ 的重心

$$\bar{x}_1^{(1)} = (0.636, 0.433)', \quad \bar{x}_2^{(1)} = (0.802, 0.672)'$$

它们与 $\bar{x}_1^{(0)}, \bar{x}_2^{(0)}$ 不同,将 $\bar{x}_1^{(1)}, \bar{x}_2^{(1)}$ 作为新的聚点,把 $x_i(i = 1, 2, \cdots, 12)$ 逐个输入,按最靠近聚点归类,其结果列在表 3.4. 得到了第二次按批修改的两个类:

$$\pi_1^{(2)} = \{x_1 \sim x_4, x_6, x_8, x_{10}\}, \pi_2^{(2)} = \{x_5, x_7, x_9, x_{11}, x_{12}\}.$$

继续进行第三次修改,这次修改后的分类结果与第二次完全相同. 分类停止. 第三次修改结果列在表 3.5.因此得到最后的分类为:

表 3.2　12 个样品点初始分类

聚点＼距离＼样品号	1	2	3	4	5	6	7	8	9	10	11	12
$x_5=(0.688,0.605)'$	0.221	0.192	0.100	0.259	0	0.274	0.217	0.307	0.145	0.152	0.055	0.148
$x_{12}=(0.777,0.723)'$	0.345	0.339	0.235	0.385	0.161	0.421	0.100	0.455	0.063	0.295	0.148	0
所属类	$\pi_1^{(0)}$	$\pi_1^{(0)}$	$\pi_1^{(0)}$	$\pi_1^{(0)}$	$\pi_1^{(0)}$	$\pi_1^{(0)}$	$\pi_2^{(0)}$	$\pi_1^{(0)}$	$\pi_2^{(0)}$	$\pi_1^{(0)}$	$\pi_1^{(0)}$	$\pi_2^{(0)}$

表 3.3　12 个样品点第一次修改分类

聚点＼距离＼样品号	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{x}_1^{(0)}=(0.647,0.450)$	0.089	0.045	0.071	0.126	0.161	0.134	0.349	0.164	0.266	0	0.161	0.302
$\bar{x}_2^{(0)}=(0.823,0.704)'$	0.339	0.351	0.237	0.377	0.167	0.435	0.055	0.468	0.032	0.305	0.148	0.045
所属类	$\pi_1^{(1)}$	$\pi_1^{(1)}$	$\pi_1^{(1)}$	$\pi_1^{(1)}$	$\pi_1^{(1)}$	$\pi_1^{(1)}$	$\pi_2^{(1)}$	$\pi_1^{(1)}$	$\pi_2^{(1)}$	$\pi_1^{(1)}$	$\pi_2^{(1)}$	$\pi_2^{(1)}$

表 3.4　12 个样品点第二次修改分类

聚点 ＼ 距离 ＼ 样品号	1	2	3	4	5	6	7	8	9	10	11	12
$x_1^{(1)}=(0.636,0.433)'$	0.084	0.032	0.089	0.122	0.179	0.114	0.369	0.148	0.302	0.032	0.179	0.322
$x_2^{(1)}=(0.802,0.672)'$	0.302	0.313	0.200	0.339	0.130	0.397	0.084	0.430	0	0.266	0.110	0.055
所属类	$\pi_1^{(2)}$	$\pi_1^{(2)}$	$\pi_1^{(2)}$	$\pi_1^{(2)}$	$\pi_2^{(2)}$	$\pi_1^{(2)}$	$\pi_2^{(2)}$	$\pi_1^{(2)}$	$\pi_2^{(2)}$	$\pi_1^{(2)}$	$\pi_2^{(2)}$	$\pi_2^{(2)}$

表 3.5　12 个样品点第三次修改分类

聚点 ＼ 距离 ＼ 样品号	1	2	3	4	5	6	7	8	9	10	11	12
$x_1^{(2)}=(0.628,0.409)'$	0.084	0.032	0.114	0.110	0.205	0.100	0.392	0.126	0.326	0.055	0.205	0.348
$x_2^{(2)}=(0.779,0.659)'$	0.283	0.288	0.176	0.322	0.105	0.371	0.110	0.405	0.045	0.243	0.084	0.063
所属类	$\pi_1^{(3)}$	$\pi_1^{(3)}$	$\pi_1^{(3)}$	$\pi_1^{(3)}$	$\pi_2^{(3)}$	$\pi_1^{(3)}$	$\pi_2^{(3)}$	$\pi_1^{(3)}$	$\pi_2^{(3)}$	$\pi_1^{(3)}$	$\pi_2^{(3)}$	$\pi_2^{(3)}$

$$\pi_1^{(3)} = \pi_1^{(2)} = \{x_1 \sim x_4, x_6, x_8, x_{10}\},$$
$$\pi_2^{(3)} = \pi_2^{(2)} = \{x_5, x_7, x_9, x_{11}, x_{12}\}.$$

2° 逐个修改法

按批修改法是将全部样品输入后形成各个类才改变聚点，而逐个修改法是每输入一个样品同时就改变该样品所进入的类的聚点，这种方法最早是 MacQueen 在 1967 年提出的，有些书称它为 K- 均值法。后来有人将它作了改进，最早的逐个修改法其步骤如下：

i）定出分类数目 K，按样品编号取前 K 个样品作为聚点（相对应图3.1 的第一框格：选择聚点）。

ii）将余下的 $n-K$ 个点逐个输入，每输入一个点将它归入到最靠近的聚点的那一类，立刻重新计算该类的重心，并将它代替原聚点（相对应于图 3.1 的第二框格：初始分类）

iii）将 n 个样品点再逐个输入，每输入一个将它归到最近聚点那一类，立刻重新计算该类重心以代替原聚点。如果 n 个点输完后所得的分类与原来分类相同则过程停止，否则再重复输入 n 个样品点（相对应于图 3.1 的第三框格：修改分类）。

后来经过改进了的逐个修改法是对于固定的 K 在修改分类的过程中可根据实际情况有所变化，太近的类可以合并，太远的类则可分离产生新类，具体步骤是：

i）定出三个数 K, C 和 $R(R > C > 0)$

ii）取前 K 个样品作为聚点，计算 K 个聚点两两之间的距离，如果最小的距离小于 C，则将这两个聚点合并，用这两点的重心作为新的聚点，再重复这个步骤，直到所有的聚点两两之间的距离都大于 C 为止。（注意此时聚点数目可能比 K 小）

iii）将余下的 $n-K$ 个样品点逐个输入，每输入一个样品点，如果该点到所有聚点的最小距离大于 R 则该点作为一个新的聚点；如果最小距离小于 R 则将该点归到距它最近的聚点的那一类，并立刻计算这一类的重心以代替原聚点。然后又重新验证聚点之

间的距离是否大于 C，如果小于 C 则用 ii）的办法将其合并，直到所有聚点之间的距离都大于 C。

iv）将 n 个样品从头至尾逐个输入，用步骤 iii）的办法归类，但稍有不同：某个样品点输入后如果分类与原来一样则重心不必计算，如果分类与原来不同，该样品所涉及的两类的重心都得重新计算。

v）按照 iv）的方法进行修改，如果相邻两次的修改分类完全相重合则过程停止，否则继续修改。

例 3.2　用逐个修改法对例 3.1 中 12 个样品点进行归类，取 $K=2$，$C=0.20$，$R=0.30$。

i）取前两个样品 $\boldsymbol{x}_1=(0.707,0.385)'$，$\boldsymbol{x}_2=(0.600,0.433)'$ 作聚点。计算 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 的距离 $d(\boldsymbol{x}_1,\boldsymbol{x}_2)=0.117<C$，因此合并 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，且计算其重心为

$$\boldsymbol{y}_1^{(1)}=\frac{1}{2}(\boldsymbol{x}_1+\boldsymbol{x}_2)=(0.654,0.409)'.$$

相对应的类记为 π_1，此时的 $\pi_1=\{\boldsymbol{x}_1,\boldsymbol{x}_2\}$。

ii）逐个输入余下的样品 $\boldsymbol{x}_3\sim\boldsymbol{x}_{12}$。例如输入 $\boldsymbol{x}_3$，即计算 $\boldsymbol{x}_3$ 与聚点 $\boldsymbol{y}_1^{(1)}$ 的距离 $d(\boldsymbol{x}_3,\ \boldsymbol{y}_1^{(1)})=0.104<R$，因此将 $\boldsymbol{x}_3$ 归入 π_1，此时 π_1 有 $\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3$ 三个样品，立刻计算其重心

$$\boldsymbol{y}_1^{(2)}=\frac{1}{3}(\boldsymbol{x}_1+\boldsymbol{x}_2+\boldsymbol{x}_3)=(0.667,0.441)'.$$

接着输入 $\boldsymbol{x}_4$，即计算 $\boldsymbol{x}_4$ 到 $\boldsymbol{y}_1^{(2)}$ 的距离 $d(\boldsymbol{x}_4,\ \boldsymbol{y}_1^{(2)})=0.110<R$，将 $\boldsymbol{x}_4$ 归入 π_1，再计算 π_1 的重心作为新的聚点，继续输入样品点，直到输入 $\boldsymbol{x}_7$ 时，它与聚点的距离为 $0.349>R$，因此 $\boldsymbol{x}_7$ 作为一聚点，另立一类，记为 π_2，此时 π_2 只有 $\boldsymbol{x}_7$。继续输入 $\boldsymbol{x}_8$，即计算 $\boldsymbol{x}_8$ 与两个聚点的距离，按最近距离归类的原则，$\boldsymbol{x}_8$ 应归入 π_1，此时 π_1 有 $\boldsymbol{x}_1\sim\boldsymbol{x}_6,\boldsymbol{x}_8$ 七个样品，计算它们的重心代替原聚点，这样一直做下去直至 12 个样品点全部输完得到初始分类为

$$\pi_1^{(0)} = \{x_1 \sim x_6, x_8, x_{10}\}, \quad \pi_2^{(0)} = \{x_7, x_9, x_{11}, x_{12}\}.$$

iii）进行修改分类。计算初始分类的重心：

$$z_1^{(0)} = \frac{1}{8}\sum_{i\in\pi_1} x_i = (0.636, 0.434)'$$

$$z_2^{(0)} = \frac{1}{4}\sum_{i\in\pi_2} x_i = (0.803, 0.673)'$$

且 $d(z_1^{(0)}, z_2^{(0)}) = 0.292 > C$。按照逐个修改法 iv）的原则将 $x_1 \sim x_{12}$ 从头至尾逐个输入得到第一次修改后的分类结果：

$$\pi_1^{(1)} = \{x_1 \sim x_4, x_6, x_8, x_{10}\}, \pi_2^{(1)} = \{x_5, x_7, x_9, x_{11}, x_{12}\}$$

计算 $\pi_1^{(1)}$，与 $\pi_2^{(1)}$ 的重心

$$z_1^{(1)} = \frac{1}{7}\sum_{i\in\pi_1^{(1)}} x_i = (0.628, 0.409)'$$

$$z_2^{(1)} = \frac{1}{5}\sum_{i\in\pi_2^{(1)}} x_i = (0.779, 0.659)'$$

且 $d(z_1^{(1)}, z_2^{(1)}) = 0.292 > C$，按照逐个修改法 iv）的原则进行第二次修改，即以 $z_1^{(1)}, z_2^{(1)}$ 为聚点，将 $x_1 \sim x_{12}$ 从头至尾再输入，得到第二次修改后的分类结果与第一次修改的分类结果完全一致，从而停止。即第一次修改的结果便是我们按逐个修改法得到最终的结果。

上面介绍的按批修改法和逐个修改法类似于计算方法的叠代法，它们是否收敛呢?在理论上是一个很重要的问题，对于 MacQueen 的 K-均值法，1967 年 MacQueen 本人给出了一个证明，而按批修改法的收敛性，Diday 在 1974 年也给出了一个证明，有兴趣的读者可参看[9]、[22]。

§8.4 降维法及图法

在二维平面上点图是人们所熟悉的一种方法，例如在回归分析中如果只有两个变量(一个因变量、一个自变量)，那么我们可以

根据 n 对观测数据在直角坐标系上点图，由散点图大致便可看出它们是否存在线性相关。因此，在聚类分析中寻找将多变量的样品点在平面上进行点图的方法长期以来为人们所关注的一个课题。在本节中我们将介绍一种非线性的映象降维法以及一种比较实用的星座图法。

（一） **非线性映象法**

设 $x_j = [x_{1j}, x_{2j}, \cdots, x_{pj}]'(j = 1,2,\cdots,n)$ 是 p 维空间中的 n 个样品点，我们希望将它映射到二维空间、设在二维空间的映射点是

$$y_j = [y_{1j}, y_{2j}]'(j = 1,2,\cdots,n).$$

并且要求这种映射尽可能保持距离不变的原则。这样就可根据映射点 $y_j(j = 1,2,\cdots,n)$ 在平面上点图进行聚类。设 d^*_{ij} 为原 p 维空间中点 x_i 与 x_j 的欧氏距离，d_{ij} 为二维空间中对应于 x_i，x_j 的点 y_i 与 y_j 的欧氏距离。非线性映象的降维原则是尽可能地使 d_{ij} 与 d^*_{ij} 一样长。为此，引人非线性换算判据或者说目标函数：

$$A = \frac{1}{\sum_{i<j} d^*_{ij}} \sum_{i<j} \frac{(d^*_{ij} - d_{ij})^2}{d^*_{ij}} = \frac{1}{\mathrm{NF}} \sum_{i<j} w_{ij}(d^*_{ij} - d_{ij})^2 \tag{4.1}$$

其中 $\mathrm{NF} = \sum_{i<j} d^*_{ij}$ 称为标准化因子，$w_{ij} = \frac{1}{d^*_{ij}}$ 称为权重系数。显然 A 是 $y_{ij}(i = 1,2, j = 1,2,\cdots,n)$ 的函数，它的含义是使原空间点之间的距离与对应的新空间点之间距离之差的平方达到极小来求得新空间点之间的几何构形，求 $\{y_{ij}\}$ 常常采用 Newton-Raphson 的迭代法。具体计算步骤如下：

1° 根据 n 个样品点 $x_j = (x_{1j}, x_{2j}, \cdots, x_{pj})'(j = 1,2,\cdots,n)$ 计算两两之间的欧氏距离 d^*_{ij}，得距离矩阵：

$$D^* = \begin{pmatrix} 0 & d^*_{12} & d^*_{13} & \cdots & d^*_{1n} \\ & 0 & d^*_{23} & \cdots & d^*_{2n} \\ & & \ddots & & \vdots \\ & & & \ddots & 0 \end{pmatrix}$$

计算 $\sum_{i<j} d^*_{ij}$，得 $\frac{1}{\mathrm{NF}}$.

2° 在二维空间中取 n 个初始点：

$$y_j^{(0)} = (y_{1j}^{(0)}, y_{2j}^{(0)})' \quad j = 1,2,\cdots,n.$$

一般初始点可采用样品在前两个主成分上的值(参看第九章)，由初始点 $y_j^{(0)}(j = 1,2,\cdots,n)$ 可算出两两之间的距离并代入(4.1)得到 $A^{(0)}$.

3° 由 $y_j^{(0)}(j = 1,2,\cdots,n)$ 及如下的迭代公式：

$$y_{ij}^{(m+1)} = y_{ij}^{(m)} - \beta\Delta_{ij}^{(m)} \tag{4.2}$$

$$(i = 1,2, j = 1,2,\cdots,n, m = 0,1,2\cdots)$$

其中 m 为迭代次数，β 为控制参数，可经验地取为 0.3 或 0.4，而

$$\Delta_{ij}^{(m)} = \frac{\partial A^{(m)}}{\partial y_{ij}^{(m)}} \Big/ \left| \frac{\partial^2 A^{(m)}}{\partial^2 y_{ij}^{(m)}} \right|,$$

这里

$$\frac{\partial A^{(m)}}{\partial y_{ij}^{(m)}} = \frac{-2}{\mathrm{NF}} \sum_{\substack{a=1 \\ a\neq j}}^{n} \left[\frac{d^*_{aj} - d_{aj}^{(m)}}{d^*_{aj} \cdot d_{aj}^{(m)}} \right] (y_{ij}^{(m)} - y_{ia}^{(m)})$$

$$\frac{\partial^2 A^{(m)}}{\partial^2 y_{ij}^{(m)}} = \frac{-2}{\mathrm{NF}} \sum_{\substack{a=1 \\ a\neq j}}^{n} \frac{1}{d^*_{aj} d_{aj}^{(m)}} \left[(d^*_{aj} - d_{aj}^{(m)}) - \frac{(y_{ij}^{(m)} - y_{ia}^{(m)})}{d_{aj}^{(m)}} \cdot \left(1 + \frac{d^*_{aj} - d_{aj}^{(m)}}{d_{aj}^{(m)}} \right) \right]$$

$$i = 1,2, j = 1,2,\cdots,n$$

可算出 $y_j^{(1)} = (y_{1j}^{(1)}, y_{2j}^{(1)})'$，$(j = 1,2,\cdots,n)$，并把它代入(4.1)计算得 $A^{(1)}$. 再由 $y_j^{(1)}$ 及迭代公式(4.2)可算出 $y_j^{(2)}$，$(j = 1, 2,$

表 4.1　六维空间的正三角形及其映到平面上的结果

点号＼坐标	六维点的坐标						二维点的坐标	
	1	2	3	4	5	6	1	2
1	0.000	0.000	0.000	0.000	0.000	0.000	0.770	1.542
2	0.370	0.640	−0.429	−0.061	−0.496	0.140	−0.067	0.996
3	0.739	1.280	−0.857	−0.122	−0.993	0.280	−0.905	0.450
4	1.109	1.920	−1.286	−0.184	−1.489	0.420	−1.743	−0.095
5	1.172	1.811	−1.225	−0.612	−0.687	0.030	−0.852	−0.548
6	1.236	1.701	−1.164	−1.041	−0.116	−0.360	0.040	−1.001
7	1.299	1.591	−1.102	−1.470	0.919	−0.750	0.932	−1.454
8	0.866	1.061	−0.735	−0.980	0.612	−0.500	0.872	−0.455
9	0.433	0.530	−0.367	−0.490	0.306	−0.250	0.842	0.543

$\cdots,n$)，并代入(4.1)计算得 $A^{(2)},\cdots$，一直迭代到第 S 次，使得对于事先给定的 $\delta>0$ 有 $A^{(S-1)}-A^{(S)}<\delta$ 就停止。最后以 $y_j^{(S)}$，($j=1,2,\cdots,n$) 作为 x_j($j=1,2,\cdots,n$) 在二维平面上的映象。

例 4.1 有六维空间上的九个点，它们均匀地分布在三条边上，现要将这九个点映到二维平面上，其数据列在表 4.1

取二维平面上的正方形为初始构形，平面上的八个初始点均匀分布在四条边上，另一个点在中心，进行迭代，当迭代到 $A=10^{-22}$ 时，y_j($j=1,2,\cdots,9$) 的坐标列在表 4.1 的最后两列中，

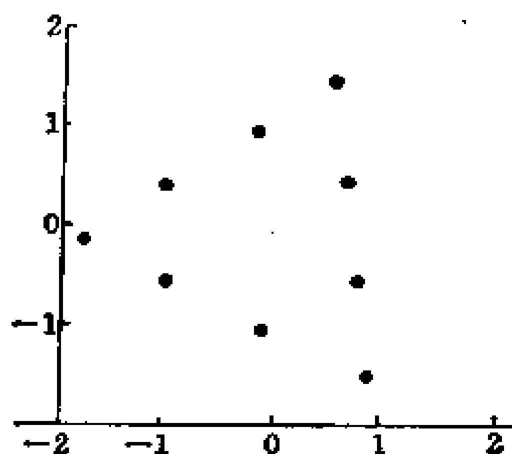


图 4.1 二维平面上的九个映象点

读者自己可去验算六维点之间的距离与对应的二维点之间的距离很接近的。将映射后的九个二维点画在平面坐标上刚好也是均匀分布在正三角形上的九个点(见图 4.1)。

降维的方法目前已提出了许多方法，如主成分法(见第九章)以及各种排序法(见[17])。

(二) 星座图

星座图是将 n 个 p 维点画在一个上半圆内，一个样品点对应一颗星，同类样品组成一个星座，不同类的样品有不同的星座，故称为星座图。

设 $x_j=(x_{1j},x_{2j},\cdots,x_{pj})'$ $j=1,2,\cdots,n$。为 n 个 p 维样品点。

先把数据 $\{x_{ij}\}$ 作一个变换，使其落在 $[0,\pi]$ 之内，这只须令

$$y_{ij}=\frac{x_{ij}-\min\limits_{1\leqslant j\leqslant n} x_{ij}}{\max\limits_{1\leqslant j\leqslant n} x_{ij}-\min\limits_{1\leqslant j\leqslant n} x_{ij}}\pi \tag{4.3}$$

相应于每一个变量 x_i 取一个权 $w_i\geqslant 0(i=1,2,\cdots,p)$，且满足 $\sum\limits_{i=1}^{p} w_i=1$。然后就可开始画星座图：

在直角坐标系中以原点 O 为圆心以单位长为半径作一个上半圆，所有的样品点按照某一规则都画在这个半圆内，每个样品按照这种规则都能找到它在半圆内相应的位置。如要找样品 x_j 的位置，先以 O 为圆心 w_1 为半径作一个上半圆，在该圆周上对应于弧度为 y_{1j} 的点记为 O_1，再以 O_1 为圆心 w_2 为半径作一个上半圆，在该圆周上对应弧度为 y_{2j} 的点记为 O_2，再以 O_2 为圆心 w_3 为半径作上半圆，在该圆周上对应于弧度 y_{3j} 的点记为 O_3，$\cdots$，如法炮制，最后在以 w_p 为半径所作的上半圆上对应着弧度 y_{pj} 的点记为 O_p。该点便是样品 x_j 的星座位置，记为 z_j。由上述的作法可看出，样品 x_j 的位置必在以单位长为半径的上半圆内。如 $p=3$ 时，取 $w_1=w_2=w_3=\frac{1}{3}$，样品点 x_j 的位置

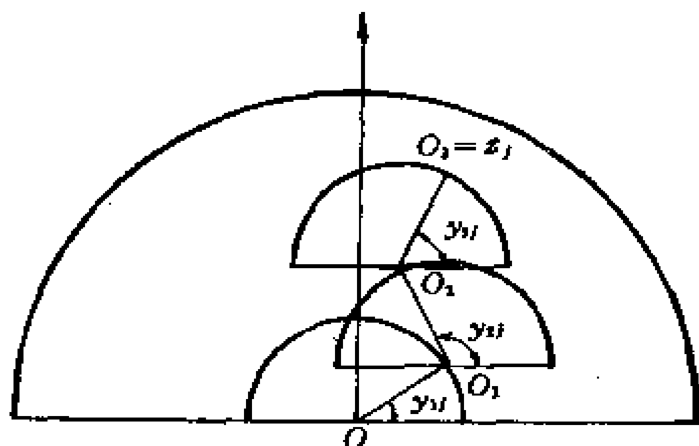


图 4.2 $p=3$，$w_i=\frac{1}{3}$，$i=1,2,3$ 时，样品 x_j 的位置 z_j

的画法如图 4.2 所示。

由 O 点出发按照上述的作图方法，x_j 达到星座位置的路线称为 x_j 的路径，路径基本上刻划了该样品的特征．在实际中我们并不是按上述的作图方法一个样品一个样品的去找它们的位置，而是直接计算它们的星座位置 z_j 的坐标，便可立刻点图了。设 $x_j(j=1,2,\cdots,n)$ 的星座位置为 $z_j(j=1,\cdots,n)$ 在直角坐标系中容易证明 z_j 的两个坐标为

$$\begin{cases} z_{1j}=\sum_{\alpha=1}^{p} w_\alpha \cos y_{\alpha j} \\ z_{2j}=\sum_{\alpha=1}^{p} w_\alpha \sin y_{\alpha j} \end{cases} \quad (j=1,2,\cdots,n) \qquad (4.4)$$

因此我们只须根据每个变量的权 w_i 及样品的 p 个变量的换算值，便可计算得 x_j 所对应于平面上的点 z_j 的坐标。

例 4.2　从某地段采集了十个土壤样品分析了五个指标：x_1 ＝燃烧损失，x_2＝可交换磷，x_3＝可提取铁，x_4＝总磷，x_5＝总氮。通过(4.3)式将原始数据变换为 $[0,\pi]$ 中的数据 y_{ij}，$(i=1,2,\cdots,5.\ j=1,2,\cdots,10)$ 并列于表 4.2，取

$$w_i=\frac{1}{5},(i=1,2,\cdots,5).$$

利用 (4.4) 式可算出 10 个土壤样品在星座图中相应的两个坐标，并画出它们相应的位置，如图 4.3. 所示．由图中可看出十个土壤样品可分为两类：

$$\{x_2\sim x_5,x_7,x_8,x_9,x_{10}\},\ \{x_1,x_6\}.$$

关于图法本世纪七十年代以来已产生了许多多变量的图表示法，如雷达图，树图，三角多项式图及连结向量图等等，由于篇幅有限我们不能一一介绍了，有兴趣的读者可参看 [24]、[25]、[29]。

表 4.2　十个土壤样品的 y_{ij}（弧度）数据

指标 \ y_{ij} \ 样号	1	2	3	4	5	6	7	8	9	10
燃烧损失	2.95	0.62	0.87	0.18	0.71	2.08	0.60	0.00	0.47	0.33
可交换磷	3.12	1.78	2.40	2.53	0.00	2.80	1.90	2.40	1.69	2.50
可提取铁	2.56	2.00	2.51	1.76	1.60	3.05	1.51	1.45	0.00	1.60
总磷	3.14	1.60	0.00	2.35	2.35	3.14	2.23	2.76	2.28	2.01
总氮	2.81	0.00	2.80	2.85	2.79	2.95	2.41	2.80	2.76	2.12

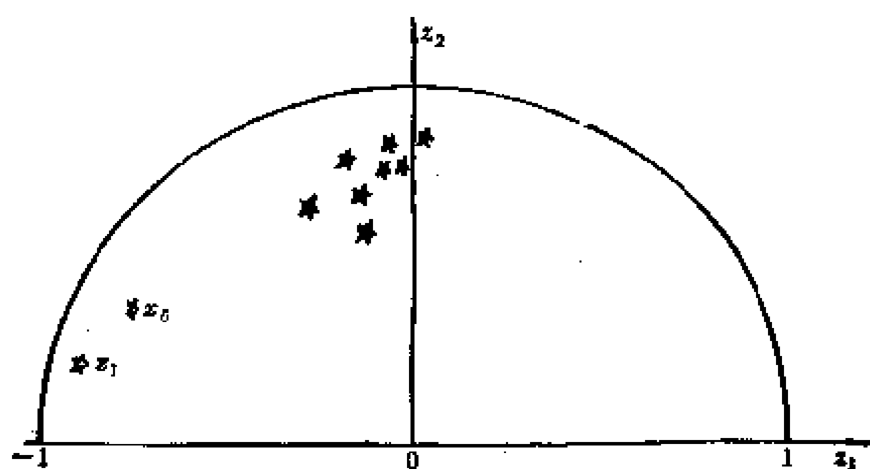


图 4.3　十个土壤样品的星座图

§ 8.5　有序样品的聚类

前面介绍的聚类方法与样品的顺序无关，但在有些实际问题中样品是有顺序的，分类不能打乱顺序。例如在地质勘探中通过钻井的岩心样品欲将地层分类，此时岩心样品是由浅到深的顺序取的。又如气象资料它是按时间排列的，其样品也是有序的。

设 $x_1, x_2, \cdots, x_n$ 是顺序取出的样品，欲将它们分成若干层（即若干类）每层的样品必须是相邻接的，即每层必须是呈现 $\{x_i;$

$x_{i+1},\cdots,x_j\}(i\geqslant 1,j\geqslant i)$ 这种形式，我们称它为样品段。

按顺序的 n 个样品，它们之间有 $n-1$ 个空隙，若在某空隙切一刀，则分割为二段；若在某两个空隙各切一刀，则分割为三段，…，若每个空隙都分别切一刀，则每一个样品各成一段。

现在对于某个 $K(1<K<n)$，要将 n 个样品分成 K 段，显然所有可能的分割有 C_{n-1}^{K-1} 种，我们要寻找出一种分割使得 K 个段的段内差异尽可能的小，各段之间的差异尽可能的大，称这种分割为相应于 K 个样品段的**最优分割**。段内的差异如何度量呢？按多元方差分析的思想，样品段 $\{x_i,x_{i+1},\cdots,x_j\}$ $(i\geqslant 1,j\geqslant i)$ 的段内差异可用广义方差来度量。

$$d_{i,j}^{*}=\left|\sum_{\beta=i}^{j}(x_\beta-\bar{x}(i,j))(x_\beta-\bar{x}(i,j))'\right| \tag{5.1}$$

其中 $\bar{x}(i,j)=\dfrac{1}{j-i+1}\sum_{\beta=i}^{j}x_\beta=(\bar{x}_1(i,j),\bar{x}_2(i,j),\cdots,\bar{x}_p(i,j))'$ 也可以用 $\sum_{\beta=i}^{j}(x_\beta-\bar{x}(i,j))(x_\beta-\bar{x}(i,j))'$ 的迹来度量，即

$$\begin{aligned}d_{i,j}&=\operatorname{tr}\left[\sum_{\beta=i}^{j}(x_\beta-\bar{x}(i,j))(x_\beta-\bar{x}(i,j))'\right]\\&=\sum_{\beta=i}^{j}\sum_{\alpha=1}^{p}(x_{\alpha\beta}-\bar{x}_\alpha(i,j))^2\end{aligned} \tag{5.2}$$

其中 $\bar{x}_\alpha(i,j)$ 为 $\bar{x}(i,j)$ 的第 α 个分量。Fisher 采用了(5.2)，因此相应的最优分割也称为 Fisher 分割。通常我们把 d_{ij} 称为样品段 $\{x_i,\cdots,x_j\}$ 的直径。

为了方便，第 i 号样品 x_i 简记为 i，把 n 个样品分割为 K 段，设某一种分割法是：

$$S(n,K):\{i_1=1,i_1+1,\cdots,i_2-1\}\{i_2,i_2+1,\cdots,i_3-1\}\cdots\{i_K,i_K+1,\cdots,n\}$$

其中 $i_1 = 1 < i_2 < \cdots < i_K \leqslant n$，定义这种分段的误差函数为

$$E(S(n,K)) = \sum_{j=1}^{K} d_{i_j, i_{j+1}-1} \tag{5.3}$$

即某种分割法的误差函数为该种分割的K个段的直径之和。当n和K固定时，$E(S(n, K))$ 越小表明在(5.2)的意义下各段内的差异越小，分类越合理，因此要寻找一种分法使得(5.3) 达到最小。设使误差函数达到最小的分法是 $S_0(n, K)$，相应误差函数值为$E(S_0(n,K))$。

1° 最优二分割

二分割是将n个样品分割为两段，一切的分法有 $C_{n-1}^1 = n-1$ 种：$\{1\}\{2,\cdots,n\},\{1,2\}\{3,\cdots,n\}\cdots,\{1,2,\cdots,n-1\}\{n\}$。显然，每种分割的误差函数为 $E(S(n, 2)) = d_{1,j-1} + d_{j,n}$ $(j = 2,\cdots,n)$，找出一种分割 $S_0(n,2)$ 使得

$$E(S_0(n,2)) = \min_{1<j\leqslant n} (d_{1,j-1} + d_{jn}) \tag{5.4}$$

则 $S_0(n,2)$ 就是最优二分割。

2° 最优三分割

三分割是将n个样品分割为三段，所有的分割法有 $C_{n-1}^2 = (n-1)(n-2)/2$ 种。每一种分割的误差函数为$E(S(n,3)) = d_{i_1,i_2-1} + d_{i_2,i_3-1} + d_{i_3,n}$（其中 $i_1 = 1$），显然，它是等价于先将n个样品分割为两段 $\{i_1 = 1, i_1+1,\cdots, i_2-1, i_2,\cdots, i_3-1\}\{i_3,\cdots,n\}$，然后再将前一段分割为两段，因此有

$$E(S(n,3)) = E(S(i_3-1,2)) + d_{i_3,n}$$

若 $S_0(n,3)$ 是最优三分割，那么应有

$$E(S_0(n,3)) = \min_{3\leqslant i_3\leqslant n} [E(S_0(i_3-1,2)) + d_{i_3,n}] \tag{5.5}$$

其中 $S_0(i_3-1,2)$ 为前 i_3-1 个样品的最优二分割。因此寻找最优三分割必须对前 i_3-1 $(3 \leqslant i_3 \leqslant n)$ 个样品进行最优二分割，然后根据(5.5)就可找到最优三分割。

3° 最优K分割

要对 n 个样品进行最优 $K(4 \leqslant K \leqslant n-1)$ 分割，与 2° 中所述的最优三分割同理等价于先将 n 个样品分割成两段：$\{i_1=1, \cdots, i_K-1\}\{i_K, \cdots, n\}$，然后将前一段再进行 $K-1$ 分割．若 $S_0(n, K)$ 是最优 K 分割，则有

$$E(S_0(n,K)) = \min_{K \leqslant i_K \leqslant n} [E(S_0(i_K-1, K-1)) + d_{i_K,n}] \quad (5.6)$$

其中 $S_0(i_K-1, K-1)$ 为前 i_K-1 个样品的最优 $K-1$ 分割．因此要寻找最优 K 分割必须对前 $i_K-1(K \leqslant i_K \leqslant n)$ 个样品先进行最优 $K-1$ 分割，然后根据 (5.6) 式就可找到最优 K 分割．

由上面的讨论可知，要进行最优 K 分割主要是计算两个表，第一个是计算 n 个样品一切可能的样品段 $\{x_i, \cdots, x_j\}(1 \leqslant i < j \leqslant n)$ 的直径 $\{d_{ij}\}$．第二个是最小误差函数表 $\{E(S_0(m, t)): t \leqslant m \leqslant n, 2 \leqslant t \leqslant K-1\}$．第一个表的计算直接按公式(5.2)．第二个表的计算是根据第一个表即直径一览表及公式 (5.4) 计算 $E(S_0(m,2)), m=2,3,\cdots, n$，然后按公式(5.5)计算 $E(S_0(m, 3)), m=3,4,\cdots,n$．以此类推一直计算到 $E(S_0(m, K-1))$，$m=K-1,K, \cdots, n$．最后根据 (5.6) 便可得最优 K 分割．注意，在具体找分割点时应该是这样：首先找出 j_K（最后一段的第一个样品号），使得

$$E(S_0(j_K-1, K-1)) + d_{j_K,n}$$

达到最小，即 j_K 是满足(5.6)式：

$$E(S_0(j_K-1,K-1)) + d_{j_K,n} = \min_{K \leqslant i_K \leqslant n} [E(S_0(i_K - 1, K-1)) + d_{i_K,n}] \quad (5.7)$$

然后再找 j_{K-1}，使它满足关系式

$$E(S_0(j_{K-1}-1, K-2)) + d_{j_{K-1},j_K} = \min_{K-1 \leqslant i_{K-1} \leqslant j_K} \cdot [E(S_0(i_{K-1}-1, K-2)) + d_{i_{K-1},j_K}] \quad (5.8)$$

以后依次类推，找出 $j_{K-2}, j_{K-3}, \cdots, j_2, j_1(=1)$，于是得到最优 K 分割为

$\{j_1=1,\cdots,j_2-1\}\{j_2,\cdots,j_3-1\}\cdots\{j_K,\cdots,n\}.$

分割数K如何确定呢？一般情况下有两个办法：(i) 可经验地取一个K值，或者按最优二分割，三分割一直作下去直到能说明你的实际问题为止。(ii) 给定一个小正数 $\delta>0$，使得当$|E(S_0(n,K-1))-E(S_0(n,K))|<\delta$，就停止，满足此不等式的$K$，即为分割数目。

我们通过一个例子来说明上述的计算步骤。

例 5.1 根据某地云杉生长过程表算出 140 年之内每 10 年直径生长量(x_1)和每 10 年树高生长量(x_2)，其数据列于表 5.1。根据此表要对云杉的生长过程进行分类，且分成三类或者四类。

表 5.1 云杉的直径和树高每10年的平均生长量

树龄	10	20	30	40	50	60	70	80	90	100	110	120	130	140
x_1	0.2	1.4	2.8	3.0	2.8	2.6	2.7	1.5	2.4	1.4	1.8	1.5	1.7	1.0
x_2	0.8	1.5	1.7	2.2	1.6	1.8	1.5	1.0	1.5	1.3	1.2	1.1	0.1	0.7

为了方便用树龄除以 10 后作为该年龄段样品的编号，如 80 年其编号为 8 号样品。

i) 按(5.2)计算 14 个样品一切可能的样品段的直径 $d_{i,j}$，并列在表5.2。例如

$$d_{3,6}=\sum_{\beta=3}^{6}\sum_{\alpha=1}^{2}(x_{\alpha\beta}-\bar{x}_\alpha(3,6))^2$$

$$=\sum_{\beta=3}^{6}[(x_{1\beta}-\bar{x}_1(3,6))^2+(x_{2\beta}-\bar{x}_2(3,6))^2]$$

$$=0.288$$

ii) 根据表 5.2 计算最小误差函数表 5.3。

首先计算分割数 $t=2$ 这一行，当 $m=2$, $t=2$，即只有前两个样品，分成二段，每段只能是一个样品，误差函数为 0。当 $m=3$, $t=2$ 时，即前三个样品分成二段，有两种分法：$\{1\}\ \{2,3\}$

表 5.2　14 个样品一切可能的样品段的直径 $d_{i,j}$

j \ i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0													
2	0.965	0												
3	3.833	1.000	0											
4	6.160	1.780	0.145	0										
5	6.884	1.930	0.233	0.200	0									
6	7.193	1.940	0.288	0.267	0.040	0								
7	7.477	2.023	0.380	0.375	0.067	0.050	0							
8	8.224	3.409	2.227	2.156	1.448	1.213	0.845	0						
9	8.291	3.420	2.266	2.180	1.448	1.230	0.947	0.530	0					
10	8.845	4.389	3.495	3.294	2.308	1.896	1.428	0.733	0.520	0				
11	8.993	4.725	3.940	3.664	2.523	2.022	1.472	0.738	0.553	0.085	0			
12	9.396	5.391	4.734	4.349	3.004	2.366	1.682	0.816	0.695	0.107	0.050	0		
13	11.125	7.368	6.765	6.245	4.582	3.789	2.877	1.862	1.804	1.028	0.787	0.520	0	
14	12.394	8.975	8.489	7.787	5.820	4.802	3.660	2.397	2.382	1.356	1.128	0.767	0.425	0

表 5.3　最小误差函数 $E(S_c(m,t))$

分割数 t \ 样品数 m	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0.00	0.965 (2)	1.110 (2)	1.198 (2)	1.252 (2)	1.345 (2)	3.191 (2)	3.230 (2)	4.388 (1)	4.725 (1)	5.390 (1)	7.369 (1)	8.975 (1)
3		0.00	0.145 (2)	0.233 (2)	0.287 (2)	0.380 (2)	1.345 (7)	1.875 (7)	2.078 (7)	2.082 (7)	2.161 (7)	3.206 (7)	3.742 (7)
4			0.00	0.145 (4)	0.185 (4)	0.211 (4)	0.380 (7)	0.910 (7)	1.113 (7)	1.117 (7)	1.196 (7)	2.161 (12)	2.586 (12)

及{1,2}{3},由(5.4)有

$$E(S_0(3,2)) = \min_{2\leqslant j\leqslant 3}(d_{1,j-1}+d_{j,3}) = \min[(d_{1,1}+d_{2,3}),(d_{1,2}+d_{33})] = \min[(0+1.00),(0.965+0)] = 0.965$$

因此对于 $m=3,t=2$ 时，最优二分割点在第二个样品之后,并用小括号列在表 5.3 相应的位置上. 当 $m=4,t=2$ 时，即前四个样品分成二段，有三种分法：{1}{2,3,4}，{1,2}{3,4}，{1,2,3}{4}. 同样由(5.4)有

$$E(S_0(4,2)) = \min_{2\leqslant j\leqslant 4}(d_{1,j-1}+d_{j,4}) = \min[(0+1.780),(0.965+0.145),(3.833+0)] = 1.110$$

分割点在第二个样品之后,并填在表 5.3 相应的位置上. 类似地可计算 $m=5,6,\cdots,14$ 的最优二分割的 $E(S_0(m,2))$ 并填在表中相应的位置上.

接着再计算表 5.3 的第二行,即分割数 $t=3$. 当 $m=3$, $t=3$ 时,显然只有一种情况,此时 $E(S_0(3,3))=0$. 当 $m=4$, $t=3$ 时,由(5.5)有

$$E(S_0(4,3)) = \min\{[E(S_0(2,2))+d_{3,4}],[E(S_0(3,2))+d_{4,4}]\} = \min\{(0+0.145),(0.965+0)\} = 0.145$$

由此可看出前 4 个样品的最优三分割应该是 {1}{2}{3,4}，第二个分割点在第二个样品之后. 并将 0.145 及第二分割点填在表 5.3 的相应位置上.如此继续下去便可把表 5.3 的第二行全部算出.

利用公式 (5.6) 可算出表 5.3 的第三行，即当分割数 $t=4$ 时，$m=4,\cdots,14$ 的各种最优四分割.

iii) 分段. 先找 14 样品分三段的最优分割点. 即找 i_3 使得

$$E(S_0(14,3)) = E(S_0(i_3-1,2))+d_{i_3,14} = \min_{j\leqslant i_3\leqslant 14}[E(S_0(i_3-1,2))+d_{i_3,n}]$$

从表中可计算出当 $i_3-1=7$ 时，$E(S_0(7,2))+d_{8,14}=1.345+2.397=3.742=E(S_0(14,3))$，因此 $i_3-1=7$ 符合要求，这表明了最优三分割的第三段为{8,…,14}，而头两段必然是{1,

2,…,7}的最优二分割，由表可查出 $j_2-1=2$为{1,2,…,7}的最优二分割的分割点（即在第二号样品之后）。这样便得到 14 个样品点的最优三分割为：

$$\{1,2\},\ \{3\sim7\},\ \{8\sim14\}.$$

用同样的手续可将 14 个样品点进行最优 4 分割．这时 $s=4$，由表 5.3 计算出

$$E(S_0(14,4))=2.586 \qquad (j_4-1=12)$$

$$E(S_0(12,3))=2.161 \qquad (j_3-1=7)$$

$$E(S_0(7,\ 2))=1.345 \qquad (j_2-1=2)$$

故最优四分割为

$$\{1,2\},\{3\sim7\},\{8\sim12\},\{13,14\}.$$

习　题　八

8.1　下面给出七个样品两两之间的距离(欧氏)矩阵：

$$D=\begin{array}{c}\\ 1\\ 2\\ 3\\ 4\\ 5\\ 6\\ 7\end{array}\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7\\ 0 & & & & & & \\ 4 & 0 & & & & & \\ 7 & 3 & 0 & & & & \\ 12 & 8 & 5 & 0 & & & \\ 18 & 14 & 11 & 6 & 0 & & \\ 19 & 15 & 12 & 7 & 1 & 0 & \\ 21 & 17 & 14 & 9 & 3 & 2 & 0 \end{pmatrix}$$

试分别用 (a) 最小距离法，(b) 最大距离法，(c) 类平均距离法，(d) 重心距离法进行聚类，并画出谱系图．

8.2　证明二态变量的相关系数为(1.9)式．

8.3　下面给出了八个样品的两个指标数据

指标＼样号	1	2	3	4	5	6	7	8
x_1	2	4	0	1	3	5	2	3
x_2	8	7	13	12	9	11	14	16

试用一分为二的分解法进行聚类，并画出谱系图。

8.4 对8.3中的八个样品，利用按批修改法进行聚类，采用欧氏距离(取1号、3号和7号样品作聚点)。

8.5 证明最大距离法、离差平方和法以及类平均法都具有单调性。

8.6 设 A、B、C 为平面上三个点，它们之间的距离为 $d^2_{A,B}=d^2_{A,C}=1.1$，$d^2_{B,C}=1$、将 A、B、C 看成三个样品，试用这个数据说明重心距离法不具有单调性。

8.7 对某地区十个县，调查了五个指标：x_1 = 土地生产率，x_2 = 劳动生产率，x_3 = 人均收人，x_4 = 费用水平，x_5 = 农村工业比重。五个指标的权数分别为 $\omega_1=0.2$，$\omega_2=0.2$，$\omega_3=0.3$，$\omega_4=0.1$，$\omega_5=0.2$。其调查数据如下表，试用星座图聚类。

指标＼县号	1	2	3	4	5	6	7	8	9	10
x_1	325	271	202	351	185	403	232	395	465	132
x_2	381	379	255	499	257	594	261	488	541	302
x_3	202	306	165	367	109	402	153	399	356	98
x_4	0.90	0.87	0.93	0.89	0.98	0.79	0.95	0.91	0.80	0.97
x_5	0.32	0.45	0.20	0.54	0.19	0.60	0.21	0.59	0.63	0.23

8.8 设有六个有序样品，一切样品段的直径 d_{ij} 如下：

j＼i	1	2	3	4	5	6
1	0					
2	0.10	0				
3	0.10	0	0			
4	0.25	0.15	0.15	0		
5	0.50	0.40	0.40	0.25	0	
6	0.50	0.40	0.40	0.34	0.34	0

试求这六个有序样品的最优三分割。

第九章　主成分分析

§9.1　概　　述

主成分分析是将研究对象的多个相关变量(指标)化为少数几个不相关的变量的一种多元统计方法．在多变量的分析中，为了尽可能完整地搜集信息，对每个样品往往要测量许多项指标，当然这可以避免重要信息的遗漏，然而从统计的角度来看，这些变量可能存在着很强的相关性，使得分析问题增加了复杂性．因此，自然想到用少数几个不相关的综合变量来代替原来较多的相关变量的研究．而且要求这些不相关的综合变量能够反映原变量提供的大部分信息，从数学的角度来看，这就是降维的思想．我们以二维的情况来直观地说明．设每个样品测定了两个变量 x_1 和 x_2，将样品根据两个变量的数据点在平面上，如图 1.1 所示． 样品点之间的差异，显然是由于 x_1 和 x_2 的变化引起的，从图上看 x_1 与 x_2 的变化范围都相差不大，如果我们将坐标轴进行一个旋转，不难看出样品点的差异主要体现在 y_1 轴上，若 y_1 所体现的差异占了大部分，譬如 85% 以上，那么我们可将 y_2 忽略，只考虑 y_1，这样两个变量就缩减了一个，问题也就相对地简化了． 显然 y_1 与 y_2 都是 x_1 和 x_2 的线性组合，称 y_1 和 y_2 分别为变量 $(x_1, x_2)'$ 的第一主成分和第二主成分(严格的定义后面再给出)．

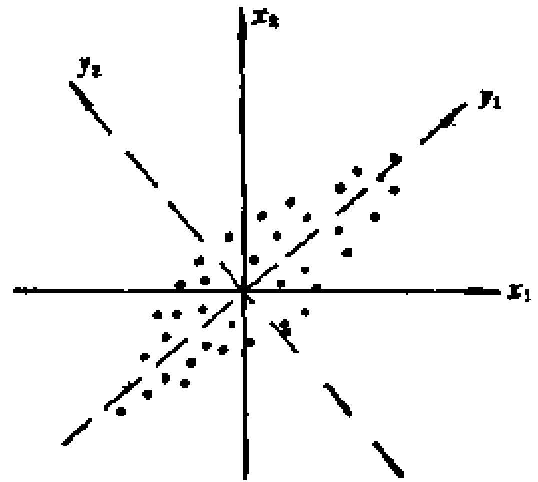


图　1.1

一般情况，设样品有 p 个变量，若样品之间的差异能由 p 个变量的 $K(K<p)$ 个主成分来概括，那么我们就能用 K 个主成分来代替 p 个变量。主成分分析往往不是目的，而是达到某种目的的一种手段，因此它常常被用在某个项目研究的中间环节，例如我们在第七章判别分析中多个总体的 Fisher 判别，建立前 r 个判别函数的思想就是主成分的思想，又如第八章聚类分析中非线性映象法的迭代中二维空间的初始点就建议采用样本在头两个主成分上的值。在本章最后一节主成分回归中，则是将主成分分析应用于回归分析，此外主成分还可用于降维的图法聚类。

从图 1.1 的直观说明，大致可以看出主成分的降维思想是从简化方差、协方差的结构来考虑的，与此种思想相一致的还有因子分析、对应分析以及典型相关分析，它们将在最后两章来讨论。

§9.2 总体的主成分

（一）未标准化变量的情况

从几何上看主成分就是把变量 $x_1,\cdots,x_p$ 构成的坐标系，在一定约束条件之下进行一个旋转产生的新坐标系中的若干个代表了具有最大变异方向的新坐标轴。从代数的观点看，主成分是 p 个变量 $x_1,\cdots,x_p$ 的一种特殊的线性组合。主成分提供了协方差阵的一种简洁的表示，后面我们将会看到主成分仅依赖于 $x_1,\cdots,x_p$ 的协方差阵（或相关系数阵），并不要求分布是正态的，对于正态总体的主成分，则借助于等密度椭球还可给出许多有用的解释。

设总体 $\boldsymbol{x}=(x_1,\cdots,x_p)'$ 的均值向量为 $\boldsymbol{\mu}$，协方差阵为 $\boldsymbol{\Sigma}$，考虑线性组合：

$$\left.\begin{aligned}
y_1 &= \boldsymbol{l}_1'\boldsymbol{x} = l_{11}x_1 + l_{21}x_2 + \cdots + l_{p1}x_p \\
y_2 &= \boldsymbol{l}_2'\boldsymbol{x} = l_{12}x_1 + l_{22}x_2 + \cdots + l_{p2}x_p \\
&\cdots\cdots\cdots\cdots\cdots \\
y_p &= \boldsymbol{l}_p'\boldsymbol{x} = l_{1p}x_1 + l_{2p}x_2 + \cdots + l_{pp}x_p
\end{aligned}\right\} \qquad (2.1)$$

由随机向量变换的性质，显然有

$$\text{var}(y_i) = \boldsymbol{l}_i'\Sigma\boldsymbol{l}_i \quad (i = 1,2,\cdots,p) \tag{2.2}$$

$$\text{cov}(y_i, y_j) = \boldsymbol{l}_i'\Sigma\boldsymbol{l}_j \quad (i,j = 1,2,\cdots,p) \tag{2.3}$$

在(2.1)中的 $y_1,y_2,\cdots,y_p$ 分别称为 $\boldsymbol{x} = (x_1,\cdots,x_p)'$ 的第一主成分，第二主成分，…，第 p 主成分，如果它们满足：

(1) y_i 与 $y_j(i \neq j)$ 不相关.

(2) y_1 是 $\boldsymbol{x}$ 的一切线性组合中方差达到最大的；y_2 是与 y_1 不相关的一切 $\boldsymbol{x}$ 的线性组合中方差达到最大的，…，y_i 是与 $y_1,\cdots,y_{i-1}(i = 2,\cdots,p)$ 都不相关的一切 $\boldsymbol{x}$ 的线性组合中方差达到最大的.

(3) $y_i = \boldsymbol{l}_i'\boldsymbol{x}$ 的系数满足 $\boldsymbol{l}_i'\boldsymbol{l}_i = 1 \quad (i = 1,2,\cdots,p)$.

上述主成分的定义中第(1)和第(2)条的含义是明确的，而第(3)条就在于不致于使 $\text{var}(y_i) \to \infty$.

下面主要是导出 $\boldsymbol{l}_i$ 及如何选择 $K(K < p)$ 个主成分使这 K 个主成分概括了 $\boldsymbol{x}$ 的变异的大部分.

定理 2.1 设总体 $\boldsymbol{x} = (x_1, x_2,\cdots,x_p)'$ 的协方差阵为 Σ，$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p \geqslant 0$ 为其特征根，$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_p$ 为对应的单位化正交特征向量，则第 i 个主成分为

$$y_i = \boldsymbol{e}_i'\boldsymbol{x} = e_{1i}x_1 + e_{2i}x_2 + \cdots + e_{pi}x_p \tag{2.4}$$
$$i = 1,2,\cdots,p$$

且有

$$\text{var}(y_i) = \boldsymbol{e}_i'\Sigma\boldsymbol{e}_i = \lambda_i \quad (i = 1,2,\cdots,p) \tag{2.5}$$

$$\text{cov}(y_i, y_j) = \boldsymbol{e}_i'\Sigma\boldsymbol{e}_j = 0 \quad (i \neq j) \tag{2.6}$$

注：若特征根有重根，对应这个特征根的特征向量组成一个 R^p 的子空间.子空间的维数等于重根数. 在子空间任取一组正交坐标系，这个坐标系的单位向量就可用来作为它的特征向量. 这时特征向量的取法不唯一，在下面的讨论中，我们总假定已选定某一种取法.

证明 由主成分的定义知求第一主成分的系数 $\boldsymbol{l}_1$ 等价于求使得 $\boldsymbol{l}'\Sigma\boldsymbol{l}$ 达到最大的 $\boldsymbol{l}_1$(在 $\boldsymbol{l}'\boldsymbol{l} = 1$ 的条件下)，根据第一章的

定理 3.14 有

$$\max_{\|l\|=1} l'\Sigma l = \lambda_1$$

且当 $l_1 = e_1$ 时达到最大,所以

$$\max_{\|l\|=1} l'\Sigma l = \lambda_1 = e_1'\Sigma e_1 = \mathrm{var}(y_1) = \mathrm{var}(e_1'x)$$

即得到 $l_1 = e_1$，且 $e_1'e_1 = 1$ 类似地有

$$\max_{\substack{l\perp e_1,\cdots,e_k \\ \|l\|=1}} l'\Sigma l = \lambda_{k+1} = e_{k+1}'\Sigma e_{k+1} = \mathrm{var}(y_{k+1}) = \mathrm{var}(e_{k+1}'x)$$

即得到 $l_{k+1} = e_{k+1}\quad (k = 1,2,\cdots,p-1)$，$e_{k+1}'e_{k+1} = 1$

由 e_i 与 $e_j(i \neq j)$ 的正交性及特征根的定义有

$$\mathrm{cov}(y_i,y_j) = e_i'\Sigma e_j = e_i'\lambda_j e_j = \lambda_j e_i'e_j = 0 \quad ▮$$

上述定理表明 x 的主成分是以 Σ 的单位化正交特征向量为系数的线性组合，第 i 个主成分的系数是对应于第 i 大特征根 λ_i 的单位化特征向量,其方差为 λ_i。

定理 2.2 设总体 $x = (x_1,\cdots,x_p)'$ 的协方差阵为 Σ，$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p \geqslant 0$ 为其特征根，相应的单位正交化特征向量为 $e_1,e_2,\cdots,e_p$，主成分为 $y_i = e_i'x(i = 1,2,\cdots,p)$。则

$$\sum_{i=1}^{p} \sigma_i^2 = \sum_{i=1}^{p} \lambda_i \tag{2.7}$$

其中 σ_i^2 为 x 的第 i 个分量 x_i 的方差，即 Σ 的对角线上第 i 个元素，λ_i 为 y_i 的方差。

证明　记 $P = [e_1,e_2,\cdots,e_p]$ 为 Σ 的特征向量矩阵,则

$$P'P = PP' = I,$$

由实对称矩阵的性质有

$$P'\Sigma P = D_\lambda \text{ 或 } \Sigma = PD_\lambda P'$$

其中 $D_\lambda = \mathrm{diag}(\lambda_1,\cdots,\lambda_p)$。

而

$$\sum_{i=1}^{p} \mathrm{var}(x_i) = \mathrm{tr}\Sigma = \mathrm{tr}(PD_\lambda P')$$

$$= \operatorname{tr}(D_\lambda P'P) = \operatorname{tr} D_\lambda = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \operatorname{var}(y_i) \quad \blacksquare$$

下面我们讨论主成分与原变量的各种相关系数之间的关系：

定理 2.3 设 $\boldsymbol{x}$ 的主成分为 $y_i = \boldsymbol{e}_i'\boldsymbol{x}, i = 1,2,\cdots,p$. 用 $\rho(y_i,x_k)$ 表示 y_i 与原变量 x_k 的相关系数，则有

$$(1)\ \rho(y_i,x_k) = e_{ki}\sqrt{\lambda_i}/\sigma_k \quad (i,k = 1,2,\cdots,p) \tag{2.8}$$

其中 e_{ki} 表示 $\boldsymbol{e}_i$ 的第 k 个分量。

$$(2)\ \sum_{k=1}^{p} \sigma_k^2\rho^2(y_i,x_k) = \lambda_i \ (i = 1,2,\cdots,p) \tag{2.9}$$

$$(3)\ \sum_{i=1}^{p} \rho^2(y_i,x_k) = 1 \ (k = 1,2,\cdots,p) \tag{2.10}$$

证明　由相关系数定义知

$$\sigma(y_i,x_k) = \frac{\operatorname{cov}(y_i,x_k)}{\sqrt{\operatorname{var}(y_i)} \cdot \sqrt{\operatorname{var}(x_k)}}$$

记 $\Phi = [\rho(y_i,x_k)]_{p\times p}$ 表示 $\boldsymbol{y} = (y_1,\cdots,y_p)'$ 与 $\boldsymbol{x} = (x_1,\cdots,x_p)'$ 的相关系数矩阵，则

$$\boldsymbol{y} = P'\boldsymbol{x} \ \text{和}\ \Phi = D_\lambda^{-\frac{1}{2}}\operatorname{cov}(\boldsymbol{y},\boldsymbol{x})D_\sigma^{-\frac{1}{2}}$$

其中 $D_\lambda = \operatorname{diag}(\lambda_1,\cdots,\lambda_p), D_\sigma = \operatorname{diag}(\sigma_1^2,\cdots,\sigma_p^2)$，而

$$\operatorname{cov}(\boldsymbol{y},\boldsymbol{x}) = P'\operatorname{cov}(\boldsymbol{x}) = P'\Sigma = (\Sigma P)' = D_\lambda P' \tag{2.11}$$

其中 $P = [\boldsymbol{e}_1,\cdots,\boldsymbol{e}_p]$ 为 Σ 的单位正交化特征向量矩阵。所以

$$\Phi = D_\lambda^{\frac{1}{2}}P'D_\sigma^{-\frac{1}{2}} \tag{2.12}$$

由(2.12)表示的 $\boldsymbol{y}$ 与 $\boldsymbol{x}$ 的相关系数矩阵的 (i,k) 元素正是(2.8)式，所以(1)得证。

记 $T = \Phi D_\sigma^{\frac{1}{2}}$，则有

$$TT' = \Phi D_\sigma\Phi' = D_\lambda^{\frac{1}{2}}P'D_\sigma^{-\frac{1}{2}}D_\sigma D_\sigma^{-\frac{1}{2}}PD_\lambda^{\frac{1}{2}} = D_\lambda^{\frac{1}{2}}P'PD_\lambda^{\frac{1}{2}} = D_\lambda \tag{2.13}$$

而 $TT' = \Phi D_\sigma\Phi'$ 的对角线的第 i 个元素正是

$$\sum_{k=1}^{p} \sigma_k^2 \rho^2(y_i, x_k),$$

由(2.13)便可看出(2.9)式成立，即(2)得证。

又由

$$\Phi'\Phi = D_\sigma^{-\frac{1}{2}} P D_\lambda^{\frac{1}{2}} D_\lambda^{\frac{1}{2}} P' D_\sigma^{-\frac{1}{2}} = D_\sigma^{-\frac{1}{2}} \Sigma D_\sigma^{-\frac{1}{2}} = R \quad (2.14)$$

上式中的R是 $\boldsymbol{x}$ 的相关系数矩阵，其对角线元素为 1. 而 $\Phi'\Phi$ 的第 k 个对角线元素为 $\sum_{i=1}^{p} \rho^2(y_i, x_k)$，由(2.14) 便可看出(2.10)成立，即(3)得证。▌

利用主成分的主要目的是降维，所以一般情况下要选取 K ($K < p$) 个主成分来代替原 p 个相关变量的研究，K 应取多大呢？在概述中我们已提到K个主成分要能概括原 p 个变量所提供信息的大部分. 为了说明各主成分 y_i 概括原变量信息的大小，我们通常用

$$a_i = \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}$$

来表示，并称它为主成分 y_i 的方差贡献率，而称 $\sum_{i=1}^{K} a_i$ 为前K个主成分的累计方差贡献率. 因此主成分个数K的大小就由累计贡献率来确定，一般情况下取K使得累计贡献率达到 80% 以上，即 $\left(\sum_{i=1}^{K} a_i\right)100\% \geqslant 80\%$.

累计贡献率是表达了K个主成分概括了原变量 $x_1, \cdots, x_p$ 的多少信息，但它没有表达某个变量 x_i 被概括了多少信息. 为此我们引进另一个概念：K个主成分 $y_1, \cdots, y_K$ 对原变量 x_i 的贡献率，这个贡献率是用 x_i 与 $(y_1, \cdots, y_K)$ 的复相关系数的平方来衡量并把它记为 τ_i，如何计算 τ_i 呢？我们有如下定理：

定理 2.4 设 $(y_1, \cdots, y_K)$ 为 $\boldsymbol{x} = (x_1, \cdots, x_p)'$ 的前K个

主成分，则 x_j 与 $(y_1,\cdots,y_K)$ 的复相关系数的平方为

$$\tau_j = \rho^2[x_j,(y_1,\cdots,y_K)] = \sum_{i=1}^{K} \lambda_i e_{ji}^2/\sigma_j^2 = \sum_{i=1}^{K} \rho^2(y_i,x_j) \quad (2.15)$$

证明　考虑向量 $(x_j,\tilde{y}')'$，其中 $\tilde{y}'=(y_1,\cdots,y_K)$ 是 $y'=(y_1,\cdots,y_p)$ 的K维子向量，于是有

$$\text{cov}(\tilde{y}) = \text{diag}(\lambda_1,\cdots,\lambda_K)$$

由(2.11)式得到

$$\alpha = \text{cov}(x_j,\tilde{y}) = (\lambda_1 e_{j1},\lambda_2 e_{j2},\cdots,\lambda_K e_{jK})'$$

由第二章复相关的公式求得

$$\rho^2[x_j,(y_1,\cdots,y_K)] = \alpha'[\text{cov}(\tilde{y})]^{-1}\alpha/\sigma_j^2$$

$$= \sum_{i=1}^{K} \lambda_i e_{ji}^2/\sigma_j^2$$

这就证明了(2.15)式的第一个等号，第二等号直接由(2.8)式立即可得. ▮

在实际应用中为了方便于对主成分的解释往往把原始变量 x_j 与前K个主成分 y_i 之间的相关系数以及复相关系数列成一个表(见表 2.1)。

如果总体 $x=(x_1,\cdots,x_p)'\sim N_p(\mu,\Sigma)(\Sigma>0)$ 我们来讨论它的主成分. 正态总体的等密度椭球为 $(x-\mu)'\Sigma^{-1}(x-\mu)=c^2$，其中 c^2 为常数，其主轴为$\pm c\sqrt{\lambda_i}\,e_i, i=1,2,\cdots,p$. 这里 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p>0$ 为Σ的特征根，$e_1,e_2,\cdots,e_p$ 为相应的单位正交化特征向量. 为了方便不妨设 $\mu=0$，记 $P=(e_1,\cdots,e_p)$，$D_\lambda=\text{diag}(\lambda_1,\cdots\lambda_p)$. 从而有

$$\Sigma^{-1} = PD_\lambda^{-1}P' = \sum_{i=1}^{p}\frac{1}{\lambda_i}e_ie_i'$$

于是

$$c^2 = x'\Sigma^{-1}x = \sum_{i=1}^{p}\frac{1}{\lambda_i}x'e_ie_i'x = \sum_{i=1}^{p}\frac{1}{\lambda_i}(e_i'x)^2$$

表 2.1　主成分与原变量的相关系数表

	$\rho(x_j, y_i)$	主成分				$\rho^2[x_j,(y_1,\cdots,y_K)]$
		y_1	y_2	……	y_K	
原始变量	x_1	$\lambda_1^{\frac{1}{2}} e_{11} \sigma_1^{-1}$	$\lambda_2^{\frac{1}{2}} e_{12} \sigma_1^{-1}$	……	$\lambda_K^{\frac{1}{2}} e_{1K} \sigma_1^{-1}$	$\frac{1}{\sigma_1^2}\sum_{i=1}^{K}\lambda_i e_{1i}^2$
	x_2	$\lambda_1^{\frac{1}{2}} e_{21} \sigma_2^{-1}$	$\lambda_2^{\frac{1}{2}} e_{22} \sigma_2^{-1}$	……	$\lambda_K^{\frac{1}{2}} e_{2K} \sigma_2^{-1}$	$\frac{1}{\sigma_2^2}\sum_{i=1}^{K}\lambda_i e_{2i}^2$
	⋮	⋮	⋮	⋮	⋮	⋮
	x_p	$\lambda_1^{\frac{1}{2}} e_{p1} \sigma_p^{-1}$	$\lambda_2^{\frac{1}{2}} e_{p2} \sigma_p^{-1}$	……	$\lambda_K^{\frac{1}{2}} e_{pK} \sigma_p^{-1}$	$\frac{1}{\sigma_p^2}\sum_{i=1}^{K}\lambda_i e_{pi}^2$

而 $e_i'x,(i=1,2,\cdots,p)$ 是 x 的主成分 y_i，所以有

$$\sum_{i=1}^{p}\frac{1}{\lambda_i}y_i^2=c^2 \tag{2.16}$$

因为 $\lambda_i>0,i=1,2,\cdots,p$. (2.16)是以主成分$y_i,i=1,2,\cdots,p$ 为主轴的一个椭球，其方向分别为 $e_i,i=1,2,\cdots,p$. 因此对于正态总体，等密度椭球的主轴方向就是主成分.

(二) 已标准化变量的情况

设总体变量 x 进行标准化，即

$$x^*=D_\sigma^{-\frac{1}{2}}(x-\mu) \tag{2.17}$$

其中 $D_\sigma=\mathrm{diag}(\sigma_1^2,\cdots,\sigma_p^2)$，$\mu$ 为 x 的均值向量. 显然 x^* 的均值向量为 $\mathbf{0}$ 向量，且 x^* 的协方差阵就是 x 的相关矩阵 R. x^* 的主成分能从 $\mathrm{cov}(x^*)=R$ 的特征向量获得，且在(一)中的结论都成立. 所不同的是，由于 $\mathrm{var}(x_i^*)=1,i=1,2,\cdots,p$. 前面定理的形式变得简单一些.

定理 2.5 对于标准化变量 $x^*=(x_1^*,\cdots,x_p^*)'$，其协方差阵为 $\mathrm{cov}(x^*)=R$，它的特征根为 $\lambda_1^*\geqslant\lambda_2^*\geqslant\cdots\geqslant\lambda_p^*$. 相应的单位化特征向量为 $e_1^*,\cdots,e_p^*$. 则有

(1) x^* 的主成分为 $y_i^*=e_i^{*\prime}x^*=e_i^{*\prime}D_\sigma^{-\frac{1}{2}}(x-\mu)$

(2) $\sum_{i=1}^{p}\mathrm{var}(y_i^*)=\sum_{i=1}^{p}\mathrm{var}(x_i^*)=p,\mathrm{var}(y_i^*)=\lambda_i^*$.

(3) $\rho(y_i^*,x_k^*)=e_{ki}^*\cdot\sqrt{\lambda_i^*}$，$(i,k=1,2,\cdots,p.)$

(4) $\sum_{k=1}^{p}\rho^2(y_i^*,x_k^*)=\lambda_i^*\quad(i=1,2,\cdots,p)$

(5) $\sum_{i=1}^{p}\rho^2(y_i^*x_k^*)=1\quad(k=1,2,\cdots,p)$

(6) x_i^* 与前K个主成分的复相关系数平方为

$$\rho^2[x_i^*,(y_1^*,\cdots,y_K^*)]=\sum_{i=1}^{K}\lambda_i^* e_{ji}^{*2}.$$

例 2.1　设 $\boldsymbol{x}=(x_1,x_2)'$ 的协方差阵及相关系数矩阵分别为

$$\Sigma=\begin{bmatrix}1 & 4\\ 4 & 100\end{bmatrix},\quad R=\begin{bmatrix}1 & 0.4\\ 0.4 & 1\end{bmatrix}$$

容易计算Σ的特征根及特征向量分别为

$$\lambda_1=100.16,\quad \boldsymbol{e}_1=(0.044,0.999)'$$

$$\lambda_2=0.84,\quad \boldsymbol{e}_2=(0.999,-0.044)'$$

未标准化变量 $\boldsymbol{x}=(x_1,x_2)'$ 的两个主成分为

$$y_1=0.044x_1+0.999x_2$$

$$y_2=0.999x_1-0.044x_2$$

计算R的特征根及特征向量分别为

$$\lambda_1^*=1.4,\quad \boldsymbol{e}_1^*=(0.707,0.707)'$$

$$\lambda_2^*=0.6,\quad \boldsymbol{e}_2^*=(0.707,-0.707)'$$

标准化变量 $\boldsymbol{x}^*=(x_1^*,x_2^*)'$ 的主成分为

$$y_1^*=0.707x_1^*+0.707x_2^*=0.707\left(\frac{x_1-\mu_1}{1}\right)$$

$$+0.707\left(\frac{x_2-\mu_2}{10}\right)$$

$$=0.707(x_1-\mu_1)+0.0707(x_2-\mu_2)$$

$$y_2^*=0.707x_1^*-0.707x_2^*=0.707\left(\frac{x_1-\mu_1}{1}\right)$$

$$-0.707\left(\frac{x_2-\mu_2}{10}\right)$$

$$=0.707(x_1-\mu_1)-0.0707(x_2-\mu_2)$$

由于 x_2 的方差很大，它几乎完全决定了从Σ算出的第一主成分，该主成分的方差贡献率为 $\frac{100.16}{101}=99.2\%$。而在标准化的情况下，$x_1^*$ 与 x_2^* 的方差都是一样，它的第一主成分的方差贡献率为

$\frac{1.4}{2}=70\%$。从这个例子可以看出变量的标准化与否对主成分的影响很大。以第一主成分为例，在由 Σ 算出的主成分 y_1 分别与 x_1, x_2 的相关系数为 $\rho(y_1, x_1)=\lambda_1^{\frac{1}{2}} e_{11}\sigma_1^{-1}=0.4404$ 和 $\rho(y_1, x_2)=\lambda^{\frac{1}{2}} e_{21}\sigma_2^{-1}=0.9998$ 即第一主成分与 x_2 最密切。而 R 算出的主成分 y_1^* 与 x_1^*, x_2^* 的相关系数为 $\rho(y_1^*, x_1^*)=\rho(y_1^*, x_2^*)=0.8365$，即第一主成分与 x_1^* 和 x_2^* 一样密切。所以一般说来，当变量取值范围彼此差异很大或者说变量的量纲不同时应该考虑标准化。

在生物学、医学中许多生物量 $x_1, \cdots, x_p$ 有时是等相关的，它们有如下形式的协方差阵和相关系数阵：

$$\Sigma=\begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \cdots & \cdots & \cdots & \cdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix}, \quad R=\begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \cdots & \cdots & \cdots & \cdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

由 R 出发，对于 $\rho>0$ 不难求得（习题 9.2）第一主成分为

$$y_1^*=\frac{1}{\sqrt{p}}\sum_{i=1}^{p} x_i^*,$$

它与 p 个变量成比例，它的方差贡献率为 $\left(\rho+\frac{1-\rho}{p}\right)\times 100\%$，当 ρ 接近于 1 时或者当 ρ 很大时，这个贡献率近似为 $\rho\times 100\%$，所以当 $\rho \geqslant 0.8$ 时，我们只须取这一个主成分就足够了。

在 Seber[30] 的书中介绍了主成分的其它形式的定义。从主成分的降维思想出发，我们要用一个 $K(K<p)$ 维向量 $\boldsymbol{y}$ 来代替 p 维向量 $\boldsymbol{x}$ 的研究，而且 $\boldsymbol{y}$ 应尽可能地概括 $\boldsymbol{x}$ 的变异。这就是说，存在一个 $p\times K$ 阵 F，使得 $F\boldsymbol{y}$ 接近 $\boldsymbol{x}$。如何来度量 $F\boldsymbol{y}$ 与 $\boldsymbol{x}$ 接近程度呢？我们用 $\boldsymbol{x}-F\boldsymbol{y}$ 的协方差阵 $V=\operatorname{cov}[(\boldsymbol{x}-F\boldsymbol{y})]$ 的一个函数 $f(V)$ 来衡量。$f(\cdot)$ 是定义在一切 $p\times p$ 阶非负定集合 $\mathscr{F}$ 上的一个函数且要满足：

i) 若 A, B 均属于 $\mathscr{F}$，且 $A\geqslant B$, $A\neq B$，则 $f(A)>$

$f(B)$,

ii) 对 $\forall A \in \mathscr{F}$, P 为任意的 p 阶正交阵，则 $f(P'AP) = f(A)$.

满足上述条件的函数$f(\cdot)$是存在的，例如$f_1(A)=\text{tr}A$, $f_2(A)=[\text{tr}AA']^{\frac{1}{2}}$, $A \in \mathscr{F}$. 设 p 维向量 $\boldsymbol{x}$ 的协方差阵 Σ 的特征根为 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p \geqslant 0$，相应的特征向量为 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_p$. 令 $\boldsymbol{P} = [\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_p] = [\boldsymbol{P}_1, \boldsymbol{P}_2]$，其中 $\boldsymbol{P}_1 = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_K]$，即 $\boldsymbol{P}$ 的前 K 列组成的 $p \times K$ 阶矩阵. 在[30]中证明了 $f(V)$ 当 $\boldsymbol{F}\boldsymbol{y} = \boldsymbol{P}_1\boldsymbol{P}_1'\boldsymbol{x}$ 时达到极小.

§9.3 样本主成分及应用实例

(一) 样本的主成分

上一节我们从总体协方差阵 Σ 或从相关系数矩阵 R 导出相应的主成分，并且不需要总体分布的假设. 当总体的协方差阵 Σ（或 R）未知时，就必须由样本 $\boldsymbol{x}_j (j=1,2,\cdots,n)$ 的协方差阵 $S = \frac{A}{n} = \frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})'$（或 $\hat{R}$）出发求样本的主成分.

对于任一线性组合 $\boldsymbol{l}_i'\boldsymbol{x}$，相应于容量为 n 的样本 $\boldsymbol{x}_j (j=1, 2,\cdots,n)$ 有

$$\boldsymbol{l}_i'\boldsymbol{x}_j = l_{1i}x_{1j} + l_{2i}x_{2j} + \cdots + l_{pi}x_{pj}, \quad (j=1,2,\cdots,n)$$

其样本均值和样本方差分别为

$$\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{l}_i'\boldsymbol{x}_j = \boldsymbol{l}_i' \cdot \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j = \boldsymbol{l}_i'\bar{\boldsymbol{x}}$$

$$\frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{l}_i'\boldsymbol{x}_j - \boldsymbol{l}_i'\bar{\boldsymbol{x}})^2 = \boldsymbol{l}_i'\left[\frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})'\right]\boldsymbol{l}_i$$

$$= \boldsymbol{l}_i'S\boldsymbol{l}_i$$

任意的两个线性组合 $\boldsymbol{l}_i'\boldsymbol{x}$ 及 $\boldsymbol{l}_k'\boldsymbol{x}$，对应于 $\boldsymbol{x}_j (j=1,\cdots,n)$ 的

样本协方差为

$$\frac{1}{n}\sum_{j=1}^{n}(l_i'x_j - l_i'\bar{x})(l_k'x_j - l_k'\bar{x})' = l_i'Sl_k$$

类似于总体主成分的定义，可以定义样本的主成分如下：

样本第一主成分 $\hat{y}_1 = l_1'x$ 是满足 $l_1'l_1 = 1$，且使得 $l_1'x_j(j=1,2,\cdots,n)$ 的样本方差 $l_1'Sl_1$ 最大的 x 的线性组合。样本的第二主成分 $\hat{y}_2 = l_2'x$ 是满足 $l_2'l_2 = 1$ 及 $l_1'x_j$ 与$l_2'x_j(j=1,2,\cdots,n)$ 的样本协方差 $l_1'Sl_2 = 0$，而且使得 $l_2'x_j(j=1,2,\cdots,n)$ 的样本方差 $l_2'Sl_2$ 最大的 x 的线性组合，…。一般地，样本的第 i 个主成分 $\hat{y}_i = l_i'x$ 是满足 $l_i'l_i = 1$ 及 $l_k'x_j$ 与 $l_i'x_j$ $(k<i, j=1,\cdots,n)$ 的样本协方差 $l_k'Sl_i = 0(k<i)$，且使得 $l_i'x_j(j=1,2,\cdots,n)$ 的样本方差 $l_i'Sl_i$ 最大的 x 的线性组合。

按照上面样本主成分的定义，容易看出也有相应于上节的定理 2.1～定理 2.4。只不过是用样本协方差阵 S 代替了总体的协方差阵 Σ 罢了，样本相关系数代替了总体相关系数，样本主成分代替总体主成分而已。

如果样本已标准化了，即 $x_j^* = D_s^{-\frac{1}{2}}(x_j - \bar{x})$，$(j=1,2,\cdots,n)$，其中 $\bar{x} = \frac{1}{n}\sum_{j=1}^{n}x_j$，$D_s = \mathrm{diag}(s_{11}, s_{22}, \cdots, s_{pp})$，$s_{ii}$ 为 $x_j(j=1,2,\cdots,n)$ 样本方差阵的对角线元素。记 $X^* = (x_1^*, x_2^*, \cdots, x_n^*)$ 表示 n 个样品组成的标准化资料阵，$\mathbf{1} = (1,1,\cdots,1)'$，那么标准化了的样本均值向量为

$$\bar{x}^* = \frac{1}{n}X^*\mathbf{1} = 0$$

样本的协方差阵为

$$S^* = \frac{1}{n}\left(X^* - \frac{1}{n}X^*\mathbf{1}\mathbf{1}'\right)\left(X^* - \frac{1}{n}X^*\mathbf{1}\mathbf{1}'\right)'$$

$$= \frac{1}{n}(X^* - \bar{x}^*\mathbf{1}')(X^* - \bar{x}^*\mathbf{1}')' = \frac{1}{n}X^*X^{*\prime} = R$$

即标准化的样本协方差阵就是样本相关系数矩阵，因此在上面样本主成分的定义中以相关系数矩阵 $\hat{R}$ 代替 S 便可得到标准化的样本主成分定义。相应地也有上一节的定理 2.5，只不过是以 $\hat{R}$ 代替总体的相关系数阵 R 罢了。

当样本中各变量之间量纲不同，即取值范围彼此差异很大时，就应该考虑将样本数据先标准化再计算样本主成分。

（二） 总体主成分的估计及其有关检验

上一段中我们是用样本协方差阵 S（或相关阵 $\hat{R}$）代替总体的协方差阵 Σ（或相关阵 R）求特征根与特征向量，并获得总体主成分 y_i 的估计 $\hat{y}_i$。自然要问 S 的特征根及特征向量是否为在某种意义下 Σ 的特征根及特征向量的估计？我们将以正态总体的情况来讨论此问题。

设 $x_1,\cdots,x_n$ 是从 $N_p(\mu,\Sigma)$ 中抽取的容量为 n 的样本，$(\Sigma>0,n>p)$，记 $\bar{x}=\frac{1}{n}\sum_{j=1}^{n}x_j$，$A=\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'$

则 μ 和 Σ 的最大似然估计是 $\bar{x}$ 和 $\frac{1}{n}A$。

Okanoto 在 1973 年证明了当 $n>p$ 时，A 以概率 1 正定且 A 的诸特征根以概率 1 互不相同．设 $S=\frac{1}{n}A$ 的特征根为 $\nu_1>\nu_2>\cdots>\nu_p>0$，相应的单位化特征向量为 $\gamma_1,\gamma_2,\cdots,\gamma_p$，记 $\Gamma=(\gamma_1,\gamma_2,\cdots,\gamma_p)$，$D_\nu=\mathrm{diag}(\nu_1,\cdots,\nu_p)$．由于 γ_i 和 $-\gamma_i$ 都是 ν_i 相应的特征向量，为了确定性，我们要求 $\gamma_i(i=1,2,\cdots,p)$ 的各分量中第一个不为 0 的分量取为“+”，那么 γ_i 就唯一确定了，从而 Γ 也唯一确定。

定理 3.1 设 $x_1,\cdots,x_n$ 是总体 $N_p(\mu,\Sigma)$ $(\Sigma>0,n>p)$ 的样本，且 Σ 的诸特征根互不相同，设 $\lambda_1>\lambda_2>\cdots>\lambda_p>0$ 为其特征根，相应的单位化特征向量为 $e_1,\cdots,e_p$ 且它们的第一个不为 0 的分量为“+”，则 λ_i 和 $e_i(i=1,2,\cdots,p)$的极大似然估计

分别为 $S=\frac{1}{n}A$ 的特征根 ν_i 及特征向量 $\boldsymbol{\gamma}_i(i=1,2,\cdots,p)$.

证明　由于 Σ 的诸特征根 λ_i 互不相同，且相应的单位化特征向量 $\boldsymbol{e}_i$ 的第一个不为 0 的分量为"+",记 $P=(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_p)'$, $D_\lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_p)$, 从而 D_λ 和 P 由 Σ 唯一确定,即 D_λ 和 P 是 Σ 的单值函数,而 $S=\frac{1}{n}A$ 是 Σ 的极大似然估计，再由第四章引理 2.1,即极大似然估计不变性知道 S 的相应特征根 ν_i 与特征向量 $\boldsymbol{\gamma}_i$ 是 λ_i 及 $\boldsymbol{e}_i$ 的极大似然估计. ▌

由于 $\frac{1}{n}A$ 不是 Σ 的无偏估计,在实用常用 $S=\frac{1}{n-1}A$ 的特征根与特征向量作为 Σ 的特征根与特征向量的估计.

上述定理的证明是设 Σ 没有重根，若 Σ 有重根，例如 $\lambda_i=\lambda_{i+1}=\cdots=\lambda_{i+(r-1)}$ 即 r 个重根,可以证明它的极大似然估计为 $\frac{1}{r}\sum_{j=0}^{r-1}\nu_{i+j}$. 至于特征向量只能在某种意义下看成极大似然估计.

关于 ν_i 及 $\boldsymbol{\gamma}_i(i=1,2,\cdots,p)$ 的分布有下面的渐近结果,它的证明在[7][13]中有较详细的讨论.

定理 3.2　设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 是 $N_p(\boldsymbol{\mu},\Sigma)(\Sigma>0,n>p)$ 的样本,其中 $\boldsymbol{\mu},\Sigma$ 未知，$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p$ 为 Σ 的特征根,相应的特征向量为 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_p$. 记样本 $S=\frac{1}{n-1}A$ 的特征根为 $\nu_1>\nu_2>\cdots>\nu_p>0$, 相应的特征向量为 $\boldsymbol{\gamma}_i,(i=1,2,\cdots,p)$. 则有

(i) 如果 Σ 的特征根互不相等,当 $n\to\infty$ 时,$\sqrt{n-1}(\nu_i-\lambda_i)/(2\lambda_i^2)^{\frac{1}{2}}$ $(i=1,2,\cdots,p)$, 渐近地相互独立且依次渐近地服从 $N(0,1)$.即使 Σ 有重根,只要 Σ 的第 j 个特征根是单根,$\sqrt{n-1}(\nu_j-\lambda_j)/(2\lambda_j^2)^{\frac{1}{2}}$ 仍渐近地服从 $N(0,1)$.

(ii) 如果 Σ 的特征根互不相等,那么当 $n\to\infty$ 时,

$$\sqrt{n-1}(\boldsymbol{\gamma}_i-\boldsymbol{e}_i)\ (i=1,2,\cdots,p)$$

渐近地服从均值向量为 0，协方差矩阵为

$$\sum_{\substack{i=1\\ i\neq j}}^{p} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} e_i e_i'$$

的 p 维正态分布。

前面也曾讨论了主成分的个数K的取法，即前K个特征根之和应占全部特征根之和的绝大部分，如果从假设检验来提这个问题，就是要检验假设

$$\begin{aligned} &H_0:\ \lambda_{K+1} = \lambda_{K+2} = \cdots = \lambda_p = 0, \\ &H_1:\ \lambda_i(i = K+1, \cdots, p)\ \text{不全为}\ 0 \end{aligned} \tag{3.1}$$

更一般地提法是

$$\begin{aligned} &H_0:\ \lambda_{K+1} = \cdots = \lambda_p = \lambda, \\ &H_1:\ \lambda_i(i = K+1, \cdots, p)\ \text{不全相等} \end{aligned} \tag{3.2}$$

定理 3.3 设 $x_1, \cdots, x_n$ 是 $N_p(\mu, \Sigma)$ $(\Sigma > 0, n > p)$ 的样本，μ 和 Σ 未知，Σ 的特征根为 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$，样本协方差阵 $S = \frac{1}{n-1}A$ 的特征根为 $\nu_1 > \nu_2 > \cdots > \nu_p$，检验假设(3.2)的似然比统计量为

$$Q = \left[\left(\prod_{i=K+1}^{p} \nu_i\right) \Big/ \left(\sum_{i=K+1}^{p} \frac{\nu_i}{p-K}\right)^{p-K}\right]^{\frac{n}{2}}$$

此定理的证明可参看[13]。

James 1969 年[13]给出它的渐近分布，令

$$\tilde{Q} = -\left\{n - 1 - K - \frac{1}{6}[2(p-K) + 2(p-K)^{-1} + 1] + \bar{\nu}^2 \sum_{i=1}^{K} (\nu_i - \bar{\nu})^{-2}\right\} \ln Q$$

它渐近地服从自由度为 $f = \frac{1}{2}(p-K)(p-K+1) - 1$ 的 χ^2 分布，其中 $\bar{\nu} = \frac{1}{p-K}\sum_{i=K+1}^{p} \nu_i$. 对于给定显著水平 α 当 $\tilde{Q} >$

$\chi^2_{1-\alpha}(f)$ 时拒绝 H_0。

在实际应用中，样本主成分的个数K的选取仍然采用总体主成分的办法，即使得前K个样本主成分的贡献率达到 80% 以上。

（三） 实例

例 3.1　某沉积盆地有一坳陷区对十七个点进行了勘探，每一个点测定了六个地质变量： x_1 = 有机碳(%)，x_2 = 生油层埋深(米)，x_3 = 油层孔隙度(%)，x_4 = 储层厚度(米)，x_5 = 地下水含碘量（p.p.m），x_6 = 地下水矿化度(克/升)。由于地质变量的单位不一样，取值范围悬殊很大，例如有机碳取值在 0.2～1.5% 之间，生油层埋深则在 1200～3200 米之间，因此将原始数据标准化，并列在表 3.1，我们要对六个地质变量进行主成分分析，其目的有两个：①找出综合地质因素与地质变量的关系；②利用主成分将 17 个勘探点进行分类为地质类比提供依据。 下面是计算和分析步骤：

(i) 根据表 3.1 的标准化数据，计算六个变量的相关系数矩阵$\hat{R}$，列于表 3.2。

(ii) 计算相关系数矩阵 $\hat{R}$ 的特征根及相应的单位化特征向量，列于表 3.3，表中还列出特征根的累积百分比。

(iii) 从表 3.3 可看出 $\hat{R}$ 的前两个特征根的累积百分比已达 81.2%，即前两个主成分的贡献率已达 81.2%，故取两个样本主成分如下：

$$\begin{cases} \hat{y}_1 = e_1'x = 0.4907x_1 + 0.4974x_2 - 0.0951x_3 \\ \qquad + 0.0718x_4 + 0.5018x_5 - 0.4957x_6 \\ \hat{y}_2 = e_2'x = 0.1882x_1 + 0.0418x_2 + 0.6148x_3 \\ \qquad + 0.7484x_4 - 0.1282x_5 + 0.0889x_6 \end{cases} \tag{3.2'}$$

(iv) 计算两个主成分与六个原变量的相关系数及每个变量与两个主成分的复相关系数并列于表 3.4。

(v) 根据表 3.4 的数据可以看出第一主成分与 x_1（有机碳），x_2（生油层埋深），x_5（地下水含碘量），x_6（地下水矿化度）这四

表 3.1　十七个勘探点的标准化数据

点号＼变量	x_1	x_2	x_3	x_4	x_5	x_6
1	−0.9142	−0.7119	−0.9293	−0.4385	−0.5710	0.7361
2	−0.3095	−0.5206	−1.3309	−0.2764	−0.5710	0.5714
3	−1.0654	−0.7119	0.2756	−0.7626	−1.0957	0.9007
4	−1.3073	−0.9511	1.2574	0.3718	−1.0957	1.3946
5	0.1743	−0.4727	0.3203	−0.9895	−0.0463	−0.2518
6	−0.8235	−0.5923	0.4095	1.3441	−0.8333	0.4068
7	0.9000	2.1583	−0.1260	−0.8598	1.7901	−1.8983
8	−0.0071	−0.3532	−1.4201	−1.0219	−0.0463	−0.5811
9	1.2023	1.6799	−0.7508	−0.6005	2.3148	−1.2397
10	0.1743	−0.3532	−0.9739	1.3441	−0.0463	0.2421
11	2.2606	1.4407	0.7219	2.6405	0.7407	−1.0750
12	−1.4282	−0.9511	0.0079	−0.7950	−1.0957	1.0653
13	−0.3397	−0.5206	2.1499	−0.1144	−0.5710	0.4068
14	0.7790	−0.2336	1.1970	0.6959	0.2161	0.9104
15	0.4162	0.7232	1.0789	−0.3088	0.4784	0.7457
16	−0.6118	−0.7119	0.3649	0.0477	−0.5710	1.5593
17	0.9000	1.0820	0.1418	−0.2764	1.0031	−0.5811

表 3.2　六个地质变量的相关系数阵 $\hat{R}$

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.0000					
x_2	0.8402	1.0000				
x_3	−0.1139	−0.0158	1.0000			
x_4	0.3474	0.0775	0.1521	1.0000		
x_5	0.8385	0.9388	−0.1942	−0.0370	1.0000	
x_6	−0.8594	−0.8631	0.2051	−0.0286	−0.8790	1.0000

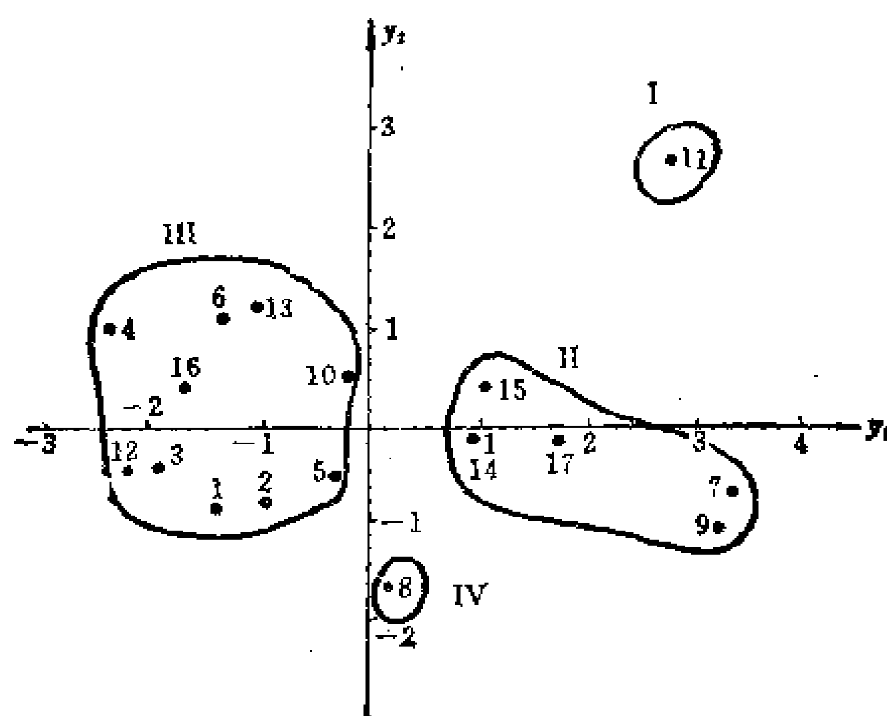


图 3.1　以两个主成分为坐标轴 17 个点的分类图

个变量特别密(其中与 x_6 呈显著的负相关)，而这些地质变量的结合大致反映了"生油条件"这个综合地质因素，因此主成分 y_1 是"生油"因素。第二主成分与 x_3（油层孔隙度），x_4（储层厚度）特别密切，这两个地质变量相结合大致反映了"储油条件"这个综合地

表 3.3 相关系数矩阵 $\hat{R}$ 的特征根及特征向量(单位化)

特征根 λ	$\lambda_1 = 3.6502$	$\lambda_2 = 1.2195$	$\lambda_3 = 0.8608$	$\lambda_4 = 0.1500$	$\lambda_5 = 0.0840$	$\lambda_6 = 0.0355$
累积百分比	60.8	81.2	95.5	98.0	99.0	100
单位特征向量	e_1	e_2	e_3	e_4	e_5	e_6
	0.4907	0.1882	−0.1388	−0.2850	−0.7447	−0.2622
	0.4974	0.0418	0.2090	0.4831	0.3277	−0.6053
	−0.0951	0.6148	0.7647	−0.1193	−0.0429	0.1099
	0.0718	0.7484	0.5797	0.1355	0.2403	0.1488
	0.5018	−0.1282	0.1223	0.4322	−0.1414	0.7142
	−0.4957	0.0889	−0.0345	0.6827	−0.5084	−0.1435

表 3.4　主成分与原变量相关系数表

$\rho(\hat{y}_i;x_j)$ 原变量 / 主成分	x_1	x_2	x_3	x_4	x_5	x_6
$\hat{y}_1$	0.9375	0.9503	−0.1817	0.1372	0.9587	−0.9471
$\hat{y}_2$	0.2078	0.0462	0.6789	0.8265	−0.1416	0.0982
$\rho^2[x_j,(\hat{y}_1,\hat{y}_2)]$	0.9221	0.9052	0.4939	0.7019	0.9391	0.9066

表 3.5　17个勘探点的主成分值

样品号 / 主成分	1	2	3	4	5	6	7	8	9
$\hat{y}_1$	−1.3973	−0.8739	−1.9542	−2.4487	−0.1495	−1.2611	3.3050	0.1473	3.2301
$\hat{y}_2$	−0.9627	−0.9812	−0.4112	1.0301	−0.5473	1.2213	−0.8599	−1.700	−1.0216

样品号 / 主成分	10	11	12	13	14	15	16	17	
$\hat{y}_1$	−0.0444	2.8515	−2.3097	−1.1264	0.9895	1.0490	−1.7452	1.7380	
$\hat{y}_2$	0.4530	2.7157	−0.6637	1.2598	−0.1868	0.4130	0.3270	−0.0855	

质因素，因此主成分 $\hat{y}_2$ 是"储油"因素．以上分析与石油地质理论是相符合的．

(vi) 要对 17 个勘探点根据两个主成分进行分类．将表 3.1 的 17 标准化样品数据分别代入主成分(3.2)式，计算出每个点的两个主成分值．计算结果见表 3.5，以 $\hat{y}_1$ 和 $\hat{y}_2$ 分别作为横坐标和纵坐标，然后将 17 个样品点在坐标系 $\hat{y}_1 O \hat{y}_2$ 上点图(见图 3.1)．从图 3.1 可看出 17 个勘探点可分为四种类型，它与实际情况基本符合．

例 3.2 Stone (1947) 利用了 Kuznets 和 Barger 所收集的美国 1922～1938 年有关国民收入与支出的 17 个要素的数据，这些要素包括纯公共支出，净增库存，股息、利息及外贸平衡等等．它们是按美元计算，但变量的标准差变化相当大，因此进行了标准化，从相关矩阵出发，对 17 个要素进行了主成分分析，并且提取了三个主成分，第一主成分的贡献率为 80.76%，第二主成分的贡献率为 10.59%，第三主成分的贡献率为 6.09%．三个主成分总的贡献率达到 97.44%．显然，这三个主成分几乎概括了 17 个经济变量的变异，因此有效维数从 17 维降到 3 维，从而达到主成分分析的目的，但如何解释这三个主成分呢？Stone 根据经济学的理论和知识，他认为变异是由三个要素组成：(1)总收入 z 或某些类似的量，(2) z 的变化率 Δz、(3) 经济发展与衰退的趋势项，该趋势项

表 3.6 三个主成分与单独可测量的 $z, \Delta z, t$ 之间的相关系数矩阵

	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	z	Δz	t
$\hat{y}_1$	1					
$\hat{y}_2$	0	1				
$\hat{y}_3$	0	0	1			
z	0.995	−0.041	0.057	1		
Δz	−0.056	0.948	−0.124	−0.102	1	
t	−0.396	−0.282	−0.836	−0.414	−0.112	1

可以取作时间 t 的线性项，因而这三个要素可以单独测量。Stone 将三个主成分 $\vartheta_1,\vartheta_2,\vartheta_3$，与 z，Δz，t 计算其相关系数，其结果列于表 3.6。

从表 3.6 中可以看出 ϑ_1 与 z，ϑ_2 与 Δz，ϑ_3 与 t 有着很大的相关关系，而 $z,\Delta z$，t 之间相关性较小，因此三个主成分实际就可解释成 $z,\Delta z$ 和 t。这三个要素。

*§ 9.4 主成分回归

（一） 问题的提出及复共线性

在回归分析中，回归系数的最小二乘估计（L.S 估计）具有许多优良性质，其中最主要的是 Gauss-Markov 定理，它保证了 L.S 估计在线性无偏估计类中的方差最小性。如果进一步假设误差服从正态分布，那么 L.S 估计还具有更多的优良性质。因此 L.S 估计一直被广泛地应用于回归系数的估计。但随着电子计算机的发展和普及，人们越来越多地处理含较多自变量的回归问题。同时在实践中也发现在一些情况下，L.S 估计并不很理想，甚至很不好。1955 年 Stein 指出了，当维数 $p>2$ 时正态均值向量的 L.S 估计的不容许性，即能够找到另一个估计在某种意义下一致优于 L.S 估计。之后人们开始寻找新的估计，其中主要有主成分估计，岭估计，Stein 估计及特征根估计。而主成分估计是根据我们前面介绍的主成分为背景引进的。本节主要就是介绍主成分估计回归系数的方法，上面提到的各种新的估计有一个共同点，就是有偏性，即估计量的数学期望都不等于待估计的参数，因此这些估计统称为有偏估计。

由上面的叙述，自然要提出在什么情况下 L.S 估计的性质才明显地变坏？为此，我们提出一个度量估计优劣的另一个标准：均方误差（Mean Square Error，简记为 MSE）。对于一般的参数估计问题，设 θ 为 p 维未知参数向量，而向量 $\hat{\theta}$ 为它的某种估计，$\hat{\theta}$ 的均方误差定义为

$$\mathrm{MSE}(\hat{\boldsymbol{\theta}}) \triangleq E\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|^2 = E(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \tag{4.1}$$

它度量了未知参数 $\boldsymbol{\theta}$ 与 $\hat{\boldsymbol{\theta}}$ 偏倚的大小，一个好的估计应该有较小的均方误差。为了更好地理解 $\mathrm{MSE}(\hat{\boldsymbol{\theta}})$ 的意义，将它进行分解

$$\begin{aligned}\mathrm{MSE}(\hat{\boldsymbol{\theta}}) &= E(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\\ &= E[(\hat{\boldsymbol{\theta}}-E\hat{\boldsymbol{\theta}})+(E\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})]'[(\hat{\boldsymbol{\theta}}-E\hat{\boldsymbol{\theta}})+(E\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})]\\ &= E(\hat{\boldsymbol{\theta}}-E\hat{\boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-E\hat{\boldsymbol{\theta}})+\|E\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|^2\\ &= \mathrm{trcov}(\hat{\boldsymbol{\theta}})+\|E\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|^2 \triangleq \Delta_1+\Delta_2 \end{aligned} \tag{4.2}$$

若 $\hat{\boldsymbol{\theta}}$ 的 p 个分量记为 $\hat{\theta}_i(i=1,\cdots,p)$。则

$$\Delta_1 = \mathrm{trcov}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{p} \mathrm{var}(\hat{\theta}_i) \tag{4.3}$$

$$\Delta_2 = \|E\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|^2 = \sum_{i=1}^{p} (E\hat{\theta}_i-\theta_i)^2 \tag{4.4}$$

由(4.3)和(4.4)可以看出它们分别是 $\hat{\boldsymbol{\theta}}$ 的各分量的方差之和以及偏差的平方和。如果 $\hat{\boldsymbol{\theta}}$ 是 $\boldsymbol{\theta}$ 的无偏估计，则 $\mathrm{MSE}(\hat{\boldsymbol{\theta}})=\Delta_1$，确实存在一些 $\boldsymbol{\theta}$ 的估计量 $\hat{\boldsymbol{\theta}}$，尽管对它而言 $\Delta_2\neq 0$，但 $\mathrm{MSE}(\hat{\boldsymbol{\theta}})$ 却较小，我们将把注意力放在这些估计上。

下面考虑第六章中已讨论过的线性回归模型

$$X = (\boldsymbol{\alpha}_0, \Theta^*)\begin{pmatrix}1'_n\\ Z^*\end{pmatrix} + \varepsilon \tag{4.5}$$

其中 $X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)$ 是 $p\times n$ 样本矩阵，Z^* 是中心化后的 $q\times n$ 矩阵，即它满足 $Z^*1_n=0$，Θ^* 是 $p\times q$ 的未知参数矩阵，$E\varepsilon=0$，$\mathrm{cov}(\mathrm{vec}\varepsilon)=\Sigma\otimes I_n$，$\mathrm{rk}(Z^*)=q$。

正如第六章第一节所讨论的：

$$\hat{\boldsymbol{\alpha}}_0 = \bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i, \quad \hat{\Theta}^* = XZ^{*\prime}\ (Z^*Z^{*\prime})^{-1}$$

是相应的 $\boldsymbol{\alpha}_0$ 和 Θ^* 的 L.S 估计。若考虑 $\hat{\Theta}^*$ 的均方误差，如通常一样可定义 $\mathrm{MSE}(\hat{\Theta}^*)=\mathrm{MSE}(\mathrm{vec}\hat{\Theta}^*)$。

类似于第六章定理 1.2 容易证明

$$\mathrm{cov}(\mathrm{vec}\hat{\Theta}^*)=\Sigma\otimes(Z^*Z^{*\prime})^{-1}$$

由于

$$E\hat{\Theta}^*=EXZ^*(Z^*Z^{*\prime})^{-1}=(\alpha_0 1_s'+\Theta^*Z^*)Z^{*\prime}(Z^*Z^{*\prime})^{-1}=\hat{\Theta}^*$$

可知 $\hat{\Theta}^*$ 是 Θ^* 的无偏估计,而 $\hat{\Theta}^*$ 的均方误差正是

$$\begin{aligned}\mathrm{MSE}(\hat{\Theta}^*)=\Delta_1=\mathrm{trcov}(\mathrm{vec}\hat{\Theta}^*)&=\mathrm{tr}[\Sigma\otimes(Z^*Z^{*\prime})^{-1}]\\&=\mathrm{tr}\Sigma\cdot\mathrm{tr}(Z^*Z^*)^{-1}\end{aligned}$$

由于 $Z^*Z^{*\prime}$ 的正定性,则存在正定阵 $P=(e_1,\cdots e_q)$ 使得

$$Z^*Z^{*\prime}=PD_\lambda P'$$

其中 $D_\lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_q)$, λ_i 是 $Z^*Z^{*\prime}$ 的第 i 个大的特征根,于是

$$\mathrm{MSE}(\hat{\theta}^*)=\mathrm{tr}\Sigma\cdot\sum_{i=1}^{q}\lambda_i^{-1}$$

如果 $Z^*Z^{*\prime}$ 中至少有一个特征根非常接近于0,那么 MSE($\hat{\theta}^*$) 变得非常大,因此从均方误差角度来看 $\hat{\Theta}^*$ 不再是 Θ^* 的一个良好估计。

如果 $Z^*Z^{*\prime}$ 至少有一个特征根接近于零,我们称设计矩阵 Z^* 呈现病态。即当设计矩阵 Z^* 呈现病态时,L.S 估计性质变坏。容易证明:当 Z^* 呈现病态时,它的行向量之间存在着近似的线性关系。事实上若记 $Z^*=(z^*_{(1)},\cdots z^*_{(q)})'$, $Z^*Z^{*\prime}$ 的第 i 个特征根 $\lambda_i\approx 0$,相应的特征向量为 e_i,于是

$$Z^*Z^{*\prime}e_i=\lambda_i e_i\approx 0.$$

由 $\|Z^{*\prime}e_i\|^2=e_i'Z^*Z^{*\prime}e_i\approx 0$ 可得

$$Z^{*\prime}e_i\approx 0$$

记 $e_i=(e_{i1},\cdots e_{iq})'$,上式可写成

$$e_{i1}z^*_{(1)}+e_{i2}z^*_{(2)}+\cdots e_{iq}z^*_{(q)}\approx 0 \tag{4.7}$$

(4.7)式说明了设计矩阵 Z^* 的行向量有近似线性关系,相应地称设计矩阵 Z^* 或者线性模型(4.5)存在着复共线性。关于复共线性的严重程度的度量已提出了一些行之有效的方法,这里介绍一种特征分析法。

设 Z^* 呈现病态，则 $Z^*Z^{*'}$ 至少有一个特征根接近于零，不妨设后面 $q-r$ 个特征根 $\lambda_{r+1},\cdots,\lambda_q$ 接近于零，相应的标准化特征向量为 $e_{r+1},\cdots,e_q$。由(4.7)式知道存在 $q-r$ 个复共线关系 $Z^{*'}e_i=0(i=r+1,\cdots,q)$。即 $Z^*Z^{*'}$ 有多少个特征根接近于零，设计矩阵 Z^* 就有多少个复共线关系式，然而 $Z^*Z^{*'}$ 的特征根接近于零是一个模糊的概念，在应用上，我们引入一个条件数

$$c=\lambda_1/\lambda_q.$$

经验地认为：若 $0<c<100$，认为不存在复共线性，若 $100\leqslant c\leqslant 1000$ 则认为存在中等或较强的复共线性，若 $c>1000$，则认为存在严重的复共线性。

（二） 回归系数的主成分估计

主成分估计是 Massy 于 1965 年提出的改善 L. S 估计的一种有偏估计，其背景是先求出原自变量的主成分，然后再求出因变量与主成分之间的回归系数。

我们考虑的回归模型仍为(4.5)。假设设计矩阵 Z^*已中心化，常数项 α_0 用 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ 来估计，记 $\lambda_i=\mathrm{ch}_i(Z^*Z^{*'})$ 为 $Z^*Z^{*'}$ 的第 i 大的特征根，相应的标准化正交特征向量为 $e_1,\cdots e_q$，记 $P=(e_1,\cdots,e_q)$。

如果把模型(4.5)中的 q 个自变量 $z_1,\cdots z_q$ 看成 q 维随机向量 z，那么设计矩阵 $Z^*=(z_1^*,\cdots,z_n^*)$ 的 n 个列向量就是随机向量 z 的 n 次观察值，由于已中心化，那么 $S=\frac{1}{n-1}Z^*Z^{*'}$ 就是 z 的样本协方差阵（注意到 $Z^*Z^{*'}$ 与 S 有相同的特征向量）。设对于 z 的 n 个观察值矩阵 Z^* 的样本主成分向量为 $g=(g_1,\cdots,g_q)'$，其中

$$g_i=e_i'z\quad(i=1,\cdots,q).\tag{4.8}$$

其中 e_i 是 S 或 $Z^*Z^{*'}$ 相应于 $\lambda_i=\mathrm{ch}_i(Z^*Z^{*'})$ 的标准化特征向量。而第 j 个观察向量 z_j^* 产生的主成分向量值为

$$g_j = \begin{pmatrix} e_1' \\ \vdots \\ e_q' \end{pmatrix} \cdot z_j^* = P'z_j^* \quad (j = 1, \cdots, n) \tag{4.9}$$

从而由设计矩阵 Z^* 产生的主成分设计矩阵可记为

$$G = P'Z^* = (g_{ij})_{q\times n} \tag{4.10}$$

注意到G仍为 $q \times n$ 阶矩阵,若令

$$\alpha = \Theta^* P \tag{4.11}$$

可将回归模型(4.5)改写成

$$X = (\alpha_0, \alpha)\begin{pmatrix} 1_n' \\ G \end{pmatrix} + \varepsilon \tag{4.12}$$

关于 ε 假设不变,上式仍为回归模型的典则形式.

如果设计矩阵 Z^* 呈现病态,那么 $Z^*Z^{*\prime}$ 的特征根至少有一个很小,不妨设后 $q-r$ 个很小,即 $\lambda_{r+1}, \cdots \lambda_q$ 近似于零. 这时候后 $q-r$ 个新变量(主成分) $g_{r+1}, \cdots, g_q$ 在 n 次试验中取值变化很小,这可如下说明:

由于 Z^* 已中心化,所以 $G1_n = P'Z^*1_n = 0$, G也已中心化了. 于是变量 g_i 在 n 次取值波动为

$$\sum_{j=1}^{n} (g_{ij} - \bar{g}_i)^2 = \sum_{j=1}^{n} g_{ij}^2 = g_{(i)}'g_{(i)} = e_i'Z^*Z^{*\prime}e_i = \lambda_i \approx 0$$
$$(i = r+1, \cdots, q)$$

其中 $G = (g_{(1)}, \cdots g_{(q)})'$, $g_{(i)}' = e_i'(z_1^*, \cdots, z_n^*) = e_i'Z^*$.
于是这些新变量的作用可合并入常数项,无碍大局我们可将 $g_{r+1}, \cdots g_q$ 从模型(4.12)中剔除.

基于上述思想,我们对 D_λ, α, G, P 作下列分块

$$D_\lambda = \begin{pmatrix} D_\lambda^{(1)} & 0 \\ 0 & D_\lambda^{(2)} \end{pmatrix}, \quad D_\lambda^{(1)} = \text{diag}(\lambda_1, \cdots \lambda_r)$$

$\alpha = (\alpha_1, \alpha_2)$, $\quad \alpha_1$ 为 $q \times r$ 阶矩阵

$P = (P_1, P_2)$, $\quad P_1 = (e_1, \cdots, e_r)$ 为 $q \times r$ 阶矩阵

$G = \begin{pmatrix} G_{(1)} \\ G_{(2)} \end{pmatrix}$, $\quad G_{(1)} = P_1'Z^*$ 为 $r \times n$ 阶矩阵.

于是(4.12)可改写为

$$X = \alpha_0 1'_n + \alpha_1 \cdot G_{(1)} + \alpha_2 \cdot G_{(2)} + \varepsilon \tag{4.13}$$

上述剔除 $g_{r+1},\cdots g_q$ 的过程也即用 $\tilde{\alpha}_2 = 0$ 估计 α_2. 于是可得到

$$X = \alpha_0 1'_n + \alpha_1 G_{(1)} + \varepsilon$$

由 $G_{(1)} 1_n = P'_1 Z^* 1_n = 0$, 即新设计矩阵 G_1 也已中心化, 故与模型(4.5)一样可得 α_1 的 L.S 估计为

$$\tilde{\alpha}_1 = X G'_{(1)} (G_{(1)} G'_{(1)})^{-1} \tag{4.14}$$

由于 $G_{(1)} = P'_1 Z^*$ 以及

$$P' Z^* Z^{*\prime} P = \begin{pmatrix} P'_1 \\ P'_2 \end{pmatrix} (Z^* Z^{*\prime}) (P_1, P_2) = \begin{pmatrix} D_\lambda^{(1)} & 0 \\ 0 & D_\lambda^{(2)} \end{pmatrix}$$

从而 $G_{(1)} G'_{(1)} = D_\lambda^{(1)}$, (4.14) 可写成

$$\tilde{\alpha}_1 = X G'_{(1)} D_\lambda^{(1)^{-1}} \tag{4.15}$$

注意到若不令 $\tilde{\alpha}_2 = 0$, 而是按通常方法求(4.13)模型中 α 的 L.S 估计,那么 $\tilde{\alpha}_1$ 不变,而 α_2 的 L.S 估计为

$$\hat{\alpha}_2 = X G'_{(2)} D_\lambda^{(2)^{-1}}.$$

可以计算得到 $\mathrm{cov}(\mathrm{vec}\hat{\alpha}_2) = \Sigma \otimes D_\lambda^{(2)^{-1}}$. $\hat{\alpha}_2$ 的每一分量的方差都非常大.主成分回归的关键在于令 $\tilde{\alpha}_2 = 0$, 使得到的 α 的估计性质有较大改善.

利用 $\tilde{\alpha}_1$ 和 $\tilde{\alpha}_2 = 0$ 代入(4.11)式得到原模型(4.15)的参数 Θ^* 的估计为

$$\tilde{\Theta}^* = \tilde{\alpha} P' = (\tilde{\alpha}_1, 0) \begin{pmatrix} P'_1 \\ P'_2 \end{pmatrix} = X G'_{(1)} D_\lambda^{(1)^{-1}} P'_1 \tag{4.16}$$

称 $\tilde{\Theta}^*$ 为 Θ^* 的主成分估计.

主成分估计有如下一些性质:

定理 4.1 设 $\hat{\Theta}^*$ 为模型(4.5)中参数 Θ^* 的 L.S 估计,$\tilde{\Theta}^*$ 为 Θ^* 的主成分估计,则

$$\tilde{\Theta}^* = \hat{\Theta}^* \cdot P_1 P'_1$$

证明

$$\tilde{\Theta}^* = XG'_{(1)}D_\lambda^{(1)^{-1}}P'_1 = XZ^{*\prime}P_1D_\lambda^{(1)^{-1}}P'_1$$
$$= \hat{\Theta}^*(Z^*Z^{*\prime})P_1 \cdot D_\lambda^{(1)^{-1}}P'_1 = \hat{\Theta}^*PD_\lambda P' \cdot P_1D_\lambda^{(1)^{-1}}P'_1$$
$$= \hat{\Theta}^*(P_1D_\lambda^{(1)}P'_1 + P_2D_\lambda^{(2)}P'_2)P_1D_\lambda^{(1)^{-1}}P'_1 = \hat{\Theta}^*P_1P'_1$$

上面最后一步利用了 $P'_1P_1 = I_r$ 和 $P'_2P_1 = 0$。 ▮

定理 4.2 记号同定理(4.1)则

$$E\tilde{\Theta}^* = \Theta^*P_1P'_1$$

即当 $r < q$ 时，主成分估计 $\tilde{\Theta}^*$ 是 Θ^* 的有偏估计。

证明：由于 L.S 估计 $\hat{\Theta}^*$ 是 Θ^* 的无偏估计，故

$$E\tilde{\Theta}^* = E\hat{\Theta}^*P_1P'_1 = \Theta^*P_1P'_1$$

当 $r < q$ 时 $P_1P'_1 \neq I_q$。故 $\tilde{\Theta}^*$ 是 Θ^* 的有偏估计。▮

定理 4.3 当设计矩阵 Z^* 呈现病态时，适当选取 r 可使得

$$\text{MSE}(\tilde{\Theta}^*) \leqslant \text{MSE}(\hat{\Theta}^*)$$

证明 记 $\tilde{\alpha} = (\tilde{\alpha}_1, 0)$，利用(4.15)

$$\tilde{\alpha} = XG'\begin{pmatrix} D_\lambda^{(1)^{-1}} & 0 \\ 0 & 0 \end{pmatrix}$$

$$\text{tr}[\text{cov}(\text{vec}\tilde{\alpha})] = \text{tr}\,\Sigma \otimes \begin{pmatrix} D_\lambda^{(1)^{-1}} & 0 \\ 0 & 0 \end{pmatrix}$$

$$= \text{tr}\,\Sigma \cdot \sum_{i=1}^{r} \lambda_i^{-1}.$$

另外

$$\|E\text{vec}\tilde{\alpha} - \text{vec}\alpha\|^2 = \text{vec}(0, \alpha_2)'\text{vec}(0, \alpha_2)$$

$$= \text{tr}(0, \alpha_2)\begin{pmatrix} 0 \\ \alpha'_2 \end{pmatrix} = \text{tr}\alpha_2\alpha'_2$$

由(4.16)

$$\text{MSE}(\tilde{\Theta}^*) = \text{MSE}(\tilde{\alpha}P') = \text{MSE}(\tilde{\alpha})$$
$$= \text{tr}[\text{cov}(\text{vec}\tilde{\alpha})] + \|E\text{vec}\tilde{\alpha} - \text{vec}\alpha\|^2$$
$$= \text{tr}\,\Sigma \cdot \sum_{i=1}^{r} \lambda_i^{-1} + \text{tr}\alpha_2\alpha'_2$$

$$= \mathrm{MSE}(\hat{\Theta}^*) + \left(\mathrm{tr}\,\alpha_2\alpha_2' - \mathrm{tr}\,\Sigma \cdot \sum_{i=r+1}^{q} \lambda_i^{-1} \right)$$

(4.17)

由于 Z^* 呈现病态，至少有一特征根接近于零，不妨设后 $q-r$ 个 λ_i 接近于零．此时 $\sum_{i=r+1}^{q} \lambda_i^{-1}$ 将非常大．可使(4.17)式第二项为负．从而对于所选的 r 有

$$\mathrm{MSE}(\tilde{\Theta}^*) \leqslant \mathrm{MSE}(\hat{\Theta}^*). \quad \blacksquare$$

定理 4.3 指出从均方误差角度出发，主成分估计 $\tilde{\Theta}^*$ 比 L.S 估计 $\hat{\Theta}^*$ 好．

关于主成分估计的其他一些最优性的讨论见[31]，[32]．

例 4.1　对某地区的某类消费品的销售量 x 进行调查，它与下面四个变量有关：z_1 居民可支配收入，z_2 该类消费品平均价格指数，z_3 社会该消费品保有量，z_4 其他消费品平均价格指数，根据历史资料(见表 4.1)，试求出该类消费品的需求函数关系．

(一) 计算自变量的相关系数矩阵 R (也即将原始数据 Z 首先中心化，得到 $Z^* = Z - \bar{Z} \cdot 1_n'$，设 S 为以 $Z^*Z^{*\prime}$ 的对角元素为对角元素的对角阵，则 $\frac{1}{n-1} S^{-\frac{1}{2}} Z^* Z^{*\prime} S^{-\frac{1}{2}} = R$．类似于主成分可以从 Σ 也可以从 R 出发讨论，本例我们从 R 出发进行讨论)．

$$R = \begin{pmatrix} 1.000 & & & \\ 0.9804 & 1.000 & & \\ 0.9883 & 0.9700 & 1.000 & \\ 0.9877 & 0.9981 & 0.9695 & 1.000 \end{pmatrix}$$

从 R 中看出自变量之间存在着高度相关．

R 的特征根分别为

$\lambda_1 = 3.9436$，$\lambda_2 = 0.0400$，$\lambda_3 = 0.0126$，$\lambda_4 = 0.0036$

表 4.1　某类消费品销售的原始数据

变量＼序号	1	2	3	4	5	6	7	8	9	10
z_1	82.9	88.0	99.9	105.3	117.7	131.0	148.2	161.8	174.2	184.7
z_2	92.0	93.0	96.0	94.0	100.0	101.0	105.0	112.0	112.0	112.0
z_3	17.1	21.3	25.1	29.0	34.0	40.0	44.0	49.0	51.0	53.0
z_4	94.0	96.0	97.0	97.0	100.0	101.0	104.0	109.0	111.0	111.0
x	8.4	9.6	10.4	11.4	12.2	14.2	15.8	17.9	19.6	20.8
$\hat{x}$	8.4	9.4	10.6	10.9	12.9	14.1	16.0	18.2	19.5	19.6

条件数 $c=\lambda_1/\lambda_4=1095>1000$，说明自变量存在严重的复共线性，其中第一个主成分的贡献率 $\lambda_1\Big/\sum_{i=1}^{4}\lambda_i$ 高达 98.6%，故我们只须取第一个主成分即可，相应于 λ_1 的特征向量为

$$e_1'=(0.5016,0.4998,0.4980,0.5007)$$

主成分变量为

$$g_1=e_1'z^*=0.5016z_1^*+0.4998z_2^*+0.4980z_3^*+0.5007z_4^*. \tag{4.18}$$

（此处 z^* 是已经标准化的）

利用(4.15)式可求得

$$\hat{\alpha}_1=XG_{(1)}'D_\lambda^{(1)^{-1}}=2.1512.$$

这里由于只取一个主成分，故 $D_\lambda^{(1)^{-1}}=\lambda_1^{-1}=(3.9436)^{-1}$. 用

$$\hat{\alpha}_0=\frac{1}{10}\sum_{i=1}^{10}x_i=14$$

估计 α_0. 因此以第一主成分 g_1 为自变量的回归方程为

$$\hat{x}=14+2.1512g_1$$

将(4.18)的主成分表达式代入上式，并将 z^* 还原成原始变量 $z=(z_1,z_2,z_3,z_4)'$，可以得到 x 与 z_1,z_2,z_3,z_4 的主成分回归方程为

$$\hat{x}=-23.4321+0.0297z_1+0.1325z_2+0.0826z_3+0.1682z_4$$

用这个回归方程对原始数据进行回报，其结果填在表 4.1 中的最后一行内。

习 题 九

9.1 设 $x=(x_1,x_2)'\sim N(0,\Sigma)$，其中

$$\Sigma=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$$

ρ 为 x_1,x_2 的相关系数，$\rho>0$。

1）试求出 $x=(x_1,x_2)'$ 的两个主成分．

2）指出主成分轴是由原坐标轴旋转了多少度而得到的．

3）指出当相关系数 ρ 为多大时，第一主成分的方差贡献率才能达到95%。

9.2 将上题推广到 p 个变量的等相关情况，即 $x=(x_1,\cdots,x_p)'$ 的相关系数矩阵为

$$R=\begin{pmatrix}1 & \rho & \rho & \cdots & \rho\\ \rho & 1 & \rho & \cdots & \rho\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ \rho & \rho & \rho & \cdots & 1\end{pmatrix}\quad 且\ 1>\rho>0$$

试导出它的第一个主成分．

9.3 设 $x=(x_1,x_2,x_3)'$ 的协方差阵为

$$\Sigma=\begin{pmatrix}\sigma^2 & \sigma^2\rho & 0\\ \sigma^2\rho & \sigma^2 & \sigma^2\rho\\ 0 & \sigma^2\rho & \sigma^2\end{pmatrix}\quad -\frac{1}{\sqrt{2}}<\rho<\frac{1}{\sqrt{2}}$$

试求主成分及每个主成分的方差贡献率．

9.4 若随机向量 $x=(x_1,\cdots,x_p)'$ 的协方差矩阵是 $\Sigma(\Sigma\geqslant0)$，随机向量 $y=(y_1,\cdots,y_p)'$ 的协方差矩阵是 $\Sigma+\sigma^2I$，其中 $\sigma^2>0$ 为常数，I 为单位阵，则 $L'x$ 是 x 的主成分的充分必要条件是 $L'y$ 是 y 的主成分，其中 L 为正交阵．

9.5 设对于标准化的样本数据 $x_j(j=1,\cdots,n)$。其样本主成分向量为 $y=(y_1,\cdots,y_p)'$ 其中 $y_i=e_i'x(i=1,\cdots,p)$，e_i 为样本协方差矩阵 S 的第 i 个大的特征根对应的标准化特征向量，求证

$$\bar{y}=\frac{1}{n}\sum_{j=1}^{n}y_j=0$$

其中 $y_j=(y_{1j},\cdots,y_{pj})'$ $y_{ij}=e_i'x_j\quad(i=1,\cdots,p,j=1,\cdots,n)$。

9.6 根据下列11组数据，利用主成分方法建立 y 与 x_1, x_2, x_3 的回归方程(取二个主成分)。

序号 \ 变量	x_1（总产值）	x_2（存储量）	x_3（总消费）	y（进口额）
1	149.3	4.2	108.1	15.9
2	161.2	4.1	114.8	16.4
3	171.5	3.1	123.2	19.0
4	175.5	3.1	126.9	19.1
5	180.8	1.1	132.1	18.8
6	190.7	2.2	137.7	20.4
7	202.1	2.1	146.0	22.7
8	212.4	5.6	154.1	26.5
9	226.1	5.0	162.3	28.1
10	231.9	5.1	164.3	27.6
11	239.0	0.7	167.6	26.3

第十章 因子分析

§ 10.1 概 述

本世纪初由 Karl Pearson 和 Chales Spearman 关于智力的定义和测量工作而开始了因子分析的近代发展。Spearman 对学生考试所得的分数做了分析，他注意到在分数之间的相关矩阵中存在一定的系统影响，表 1.1 是某学校 33 个学生 6 门功课的相关系数矩阵

表 1.1 六门功课的相关矩阵

	1	2	3	4	5	6
1.古典语	1					
2.法语	0.83	1				
3.英语	0.78	0.67	1			
4.数学	0.70	0.67	0.64	1		
5.判别	0.66	0.65	0.54	0.54	1	
6.音乐	0.63	0.57	0.51	0.51	0.4	1

表中课程是按照相关系数从上到下递减排列的。Spearman 指出，如果第 i 个变量(第 i 门功课)上的分数是由两部分组成的

$$x_i = a_i F + \varepsilon_i \tag{1.1}$$

那么各门功课相关的“效应”就可以被说明。其中 F 是对所有变量都起作用的公因子，ε_i 是对 x_i 所特有的，即每门课程的考试成绩可以看作由一个公因子(与智力相一致)和一个特殊因子之和组成。(1.1) 式是最早的最简单的因子模型。我们再来看一个例子，考虑人的五个生理指标：收缩压 (x_1)，舒张压 (x_2)，心跳间隔 (x_3)，呼吸间隔 (x_4)，舌下温度 (x_5)。从生理学的知识知道这五个指标是受植物神经的交感神经和副交感神经这两个因子的共同

影响，即这五个指标至少受到两个公共因子的作用，如果用 F_1, F_2 分别表示交感神经与副交感神经，那么可以设想变量 $\boldsymbol{x}=(x_1,\cdots,x_5)'$ 是 $\boldsymbol{F}=(F_1,F_2)'$ 的线性函数（即 F_i 对各个变量的影响是线性的），再加上其它对 x_i 有影响的因子，用函数的形式表达，即为

$$\left.\begin{array}{l} x_1 = a_{11}F_1 + a_{12}F_2 + g_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ x_5 = a_{51}F_1 + a_{52}F_2 + g_5 \end{array}\right\} \tag{1.2}$$

或者写成矩阵形式

$$\boldsymbol{x} = \boldsymbol{AF} + \boldsymbol{g} \tag{1.3}$$

如果 $g_1, g_2, \cdots, g_5$ 还可以分解出一些公共因子，则有

$$\boldsymbol{g} = \boldsymbol{MH} + \boldsymbol{\varepsilon}$$

此时

$$\boldsymbol{x} = \boldsymbol{AF} + \boldsymbol{MH} + \boldsymbol{\varepsilon} = (\boldsymbol{A}\boldsymbol{M})\begin{pmatrix}\boldsymbol{F}\\ \boldsymbol{H}\end{pmatrix} + \boldsymbol{\varepsilon}$$

因此总可以认为有如下这种形式：

$$\boldsymbol{x} = \boldsymbol{AF} + \boldsymbol{\varepsilon} \tag{1.4}$$

其中 $\boldsymbol{\varepsilon}$ 是无法再分解出公共因子的特殊因子。 这样五个生理变量之间的相关"效应"就可以通过公因子和特殊因子来说明。

一般情况，设有 p 个变量 $\boldsymbol{x}=(x_1,\cdots,x_p)'$ 它们受到 m 个公因子 $\boldsymbol{F}=(F_1,\cdots,F_m)'$ 的共同影响，且各个变量还单独地受到一个特殊因子 $\varepsilon_i(i=1,2,\cdots,p)$ 的作用，设这些影响是线性的，即可表达为

$$\left.\begin{array}{l} x_1 = a_{11}F_1 + \cdots + a_{1m}F_m + \varepsilon_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ x_p = a_{p1}F_1 + \cdots + a_{pm}F_m + \varepsilon_p \end{array}\right\} \tag{1.5}$$

或者写成矩阵形式

$$\boldsymbol{x} = \boldsymbol{AF} + \boldsymbol{\varepsilon} \tag{1.6}$$

其中 $a_{ij}(i=1,2,\cdots,p, j=1,2,\cdots,m.)$ 称为第 i 个变量在第

j 个公因子上的载荷，A 是由 a_{ij} 为元素组成的矩阵称它为载荷矩阵。同样地，p 个变量之间的相关关系的内在原因，就是要通过公因子及特殊因子来说明。模型(1.5)(或(1.6))是针对 p 个变量内部之间的相关关系，因此称它为 R 型的因子模型。

当然，我们可以把因子分析看为主成分分析的一种推广，但它与主成分分析是不同的，主成分分析只是从一个协方差阵(或相关系数阵)出发直接求得一组互不相关的主成分(它们是原变量的线性组合)，并选出前 K 个主成分来替代原变量。而因子分析则是要用假设的公因子来"解释"相关矩阵的内部的依赖结构。它们的数学表达是不一样的(比较一下第九章的(2.1)与(1.5))。在以后的讨论中读者将还会发现因子分析与主成分分析的其它不同之处。

如何从一组资料出发，分析出对变量有公共影响的公因子，在实际工作中是很有意义的，例如在地质科学中，可以利用因子分析构成在各种地质作用下化学元素的自然共生组合或其它地质因素的自然组合，从而对地质，地球化学现象进行成因分析。

在第八章聚类分析中，已定义过样品之间的相似性的度量，根据前面所述 R 型因子分析的思想，自然想到对 n 个具有某种程度相似性的样品，也希望表示成一些对各个样品有公共影响的公因子和对各个样品的特殊因子的线性组合。Imbrie (1963) 及 Benzécri (1972) 在不同的领域中首先做了这方面的工作(见[21])。设有 n 个样品 $x_1, \cdots, x_n$ 我们将它们表示成：

$$\left.\begin{array}{l} x_1 = a_{11}F_1 + \cdots + a_{1m}F_m + \varepsilon_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ x_n = a_{n1}F_1 + \cdots + a_{nm}F_m + \varepsilon_n \end{array}\right\} \quad (1.7)$$

其中 $F_j(j=1,2,\cdots,m)$ 是对每个样品都有影响的公共因子，而 $\varepsilon_i(i=1,2,\cdots,n)$ 只对样品 $x_i(i=1,2,\cdots,n)$ 起作用的特殊因子。我们要通过公因子和特殊因子来说明样品之间相似关系的内部结构。(1.7) 这种因子模型称为 Q 型因子模型。例如在地质学中，有一沉积盆地，假设有三个剥蚀区供给沉积物(如图 1.1 所

示）。可以想象，沉积盆地中的任一个样品的重矿物的百分含量都是由三个剥蚀区的重矿物所供给及各个样品的特殊因子的作用而得到的，因此沉积盆地的任一块样品 x_i 可表为

$$x_i = a_{i1}F_1 + a_{i2}F_2 + a_{i3}F_3 + \varepsilon_i \quad (i = 1, 2, \cdots, n).$$



图 1.1　有三个剥蚀区的沉积盆地示意图

Q 型因子分析的做法是在沉积盆地中采集许多样品，利用样品中重矿物的数据，去找出公因子 F_i 的数目及其载荷 $\{a_{ij}\}$，并赋予 F_i 具体解释，然后找出它们的古地理位置，这样我们就可恢复当时的古地理面貌。因此 Q 型因子分析在实际中也是十分有意义的。

无论是 R 型或 Q 型因子分析都是用少量几个公共因子来研究变量（或样品）的内部相关（或相似）结构，这些公因子可视为一些随机变量，且事先是无法观察的。从数学上来看，因子分析是不完善的，但本章是从实际工作的要求来说明一些理论上的概念和具体的方法。

§ 10.2　正交因子模型及其解

（一）　正交因子模型

在上一节中我们已经谈到因子模型分为 R 型和 Q 型，本节以及以后的讨论主要是针对 R 型，而 Q 型我们会通过例子来说明。

设 p 维随机向量 $\boldsymbol{x} = (x_1, \cdots, x_p)$ 的均值向量为 $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)'$，协方差阵为 Σ，不妨将 $\boldsymbol{x}$ 中心化，此时 R 型因子分析模型可写为

$$\boldsymbol{x} - \boldsymbol{\mu} = A\boldsymbol{F} + \boldsymbol{\varepsilon} \tag{2.1}$$

其中各个记号的解释同上节，注意到(2.1)中用了 $p+m$ 个不可观测的随机变量 $F_1,\cdots,F_m,\varepsilon_1,\cdots,\varepsilon_p$ 来表示 p 个变量 $x_1,\cdots,x_p$。这正是 R 型因子分析模型与回归模型的区别所在，因此对于(2.1)采用回归分析方法来确定载荷矩阵 A 是行不通的，如果我们对 F 及 ε 作一些假定，那么模型(2.1)就具有特定的且能够验证的协方差关系。假设

$$\left.\begin{aligned}&E(\boldsymbol{F})=0,\mathrm{cov}(\boldsymbol{F})=I\\&E(\boldsymbol{\varepsilon})=0,\mathrm{cov}(\boldsymbol{\varepsilon})=\phi=\mathrm{diag}(\phi_1^2,\cdots,\phi_p^2)\\&\mathrm{cov}(\boldsymbol{\varepsilon},\boldsymbol{F})=0\end{aligned}\right\}\tag{2.2}$$

将(2.1)及(2.2)合在一起称为**正交因子模型**（R 型），当 $\boldsymbol{F}$ 的各个分量相关时，$\mathrm{cov}(\boldsymbol{F})$ 不再为单位阵，其相应的模型称为**斜交因子模型**，我们将在§10.4中进行讨论。

由正交因子模型我们可得到

$$\begin{aligned}\Sigma=\mathrm{cov}(\boldsymbol{x})&=E[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})']\\&=E[(A\boldsymbol{F}+\boldsymbol{\varepsilon})(A\boldsymbol{F}+\boldsymbol{\varepsilon})']=E[A\boldsymbol{F}\boldsymbol{F}'A'+\boldsymbol{\varepsilon}\boldsymbol{F}'A'\\&\quad+A\boldsymbol{F}\boldsymbol{\varepsilon}'+\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']=AA'+\phi\end{aligned}\tag{2.3}$$

$$\mathrm{cov}(\boldsymbol{x},\boldsymbol{F})=E[(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{F}']=E[(A\boldsymbol{F}+\boldsymbol{\varepsilon})\boldsymbol{F}']=A\tag{2.4}$$

由(2.3)有

$$\sigma_{ii}=\mathrm{var}(x_i)=a_{i1}^2+a_{i2}^2+\cdots+a_{im}^2+\phi_i^2\qquad i=1,2,\cdots,p\tag{2.5}$$

$$\sigma_{ik}=\mathrm{cov}(x_i,x_k)=a_{i1}a_{k1}+a_{i2}a_{k2}+\cdots+a_{im}a_{km}\qquad i\neq k\tag{2.6}$$

由(2.4)有

$$\mathrm{cov}(x_i,F_j)=a_{ij},i=1,\cdots,p,\ j=1,\cdots,m.\tag{2.7}$$

从(2.5)式可以看出变量 x_i 的方差是由 m 个公因子和特殊因子提供的，称 a_{ij}^2 为第 j 个公因子对变量 x_i 的方差贡献，将 $a_{i1}^2+\cdots+a_{im}^2$ 记为 h_i^2，它表示了 m 个公因子对变量 x_i 的方差贡献总和，并称 h_i^2 为第 i 个变量的共同度，它刚好是载荷矩阵的第 i 行元素

平方和，而 ψ_i^2 为特殊因子提供的方差称为特殊度，所以(2.5)式可写成

$$\mathrm{var}(x_i) = \sigma_{ii} = h_i^2 + \psi_i^2 \tag{2.8}$$

不妨设 x_i 已标准化，此时 $\mathrm{var}(x_i) = 1$，当 $h_i^2 = 1$ 时，$\psi_i^2 = 0$。这表明 x_i 能被所有公因子的线性组合表示；当 h_i^2 接近于0时，表明公因子对 x_i 的影响不大，此时 x_i 由特殊因子来描述，由此可看出 h_i^2 反映了变量 x_i 对公因子依赖的程度。另一方面考虑某个指定的公因子 F_j 对各个变量 $x_1, \cdots, x_p$ 的影响用

$$g_j^2 = a_{1j}^2 + a_{2j}^2 + \cdots + a_{pj}^2 \tag{2.9}$$

来描述，它刚好为载荷矩阵第 j 列的元素的平方和。称 g_j^2 为公因子 F_j 对所有 p 个变量的方差贡献。显然，g_j^2 越大反映了 F_j 对所有变量的影响越大，它可作为公因子 F_j 重要性的一个度量。从(2.7)可看出 a_{ij} 表示了变量 x_i 与公因子 F_j 的相关系数。

从上面的讨论，可归纳出正交因子模型中的载荷矩阵 A 有如下的统计意义：

(1) $h_i^2 = a_{i1}^2 + \cdots + a_{im}^2$，为第 i 个变量的共同度它度量了变量 x_i 对 m 个公因子的依赖程度；

(2) $g_j^2 = a_{1j}^2 + \cdots + a_{pj}^2$，为第 j 个公因子 F_j 对所有变量的贡献，它作为公因子 F_j 重要性的一个度量。

(3) a_{ij} 是变量 x_i 与公因子 F_j 的相关系数，它的大小为实际工作中解释公因子的实际含义提供一种依据。

当 $m > 1$ 时，正交因子模型本身具有不确定性，即载荷矩阵并不唯一。设 T 为 $m \times m$ 阶的正交矩阵，(2.1)式可改写为

$$\boldsymbol{x} - \boldsymbol{\mu} = A\boldsymbol{F} + \boldsymbol{\varepsilon} = ATT'\boldsymbol{F} + \boldsymbol{\varepsilon} = A^*\boldsymbol{F}^* + \boldsymbol{\varepsilon} \tag{2.10}$$

其中 $A^* = AT, \boldsymbol{F}^* = T'\boldsymbol{F}$，且有

$$E(\boldsymbol{F}^*) = T'E(\boldsymbol{F}) = 0$$

$$\mathrm{cov}(\boldsymbol{F}^*) = T'\mathrm{cov}(\boldsymbol{F})T = T'T = I$$

$$\mathrm{cov}(\boldsymbol{\varepsilon}, \boldsymbol{F}^*) = \mathrm{cov}(\boldsymbol{\varepsilon}, T'\boldsymbol{F}) = E(\boldsymbol{\varepsilon} \cdot \boldsymbol{F}')T' = 0$$

于是 A^* 和 $\boldsymbol{F}^*$ 也满足模型，显然从变量 $x_1, \cdots, x_p$ 的观测值，

我们是不可能区分载荷阵 A 和 A^*，即公因子 $\boldsymbol{F}$ 和 $\boldsymbol{F}^* = T'\boldsymbol{F}$ 具有相同的统计性质，载荷矩阵 A^* 和 A 虽然不同，但它们可产生相同的协方差矩阵 Σ，即

$$\Sigma = AA' + \psi = A^*A^{*\prime} + \psi \tag{2.11}$$

由于 $\boldsymbol{F}^* = T'\boldsymbol{F}$，说明 $\boldsymbol{F}^*$ 是由 $\boldsymbol{F}$ 经过正交旋转得到，同样 $A^* = AT$ 也说明 A^* 是由 A 经过正交变换而得到，从而可推出对于 $\boldsymbol{F}$ 和 $\boldsymbol{F}^*$，变量 x_i 的共同度 h_i^2 是一样，即

$$h_i^2 = \sum_{j=1}^{m} a_{ij}^2 = \sum_{j=1}^{m} a_{ij}^{*2},$$

或者说共同度 h_i^2 与正交阵 T 的选择无关．关于这种不确定性在实际应用中十分重要，我们将在下一节看到这一点．

（二） 因子载荷阵的估计

由因子模型知道，我们试图用 m 个公共因子和特殊因子来研究相关矩阵的内部依赖结构或者说相关关系的内在原因．从本质上讲就是要建立统计模型(2.1)与(2.2)，使其满足方差结构(2.3)，即

$$\Sigma = AA' + \psi$$

我们知道协方差阵 Σ 的估计为样本协方差阵 S，如果 S 的非对角线元素都较小，即它们与 0 无显著差异，那么 p 个变量是互不相关的，此时特殊因子起着决定性作用，因此在这种情况，作因子分析是无意义的，如果 S 不是接近对角阵，那么可以考虑建立因子模型．从理论上如何去检验 Σ 具有结构(2.3)的问题，我们将在下一段进行讨论．

如果考虑建立因子模型，第一个要解决的问题是如何估计载荷矩阵 A 及特殊方差 $\psi_i(i = 1, 2, \cdots, p)$．目前已提出了许多方法，如极大似然法，主成分法以及主因子解等方法．在这里我们仅介绍极大似然法和主成分法．

1° 极大似然法

定理 2.1 设 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$，$(n > p)$ 是从正态总体 $N_p(\boldsymbol{\mu}, \Sigma)$，

($\Sigma > 0$) 抽取的容量为 n 的样本，其中协方差阵 Σ 具有结构：$\Sigma = AA' + \phi$，这里的 A 是 $p \times m$ 阶矩阵，$\mathrm{rk}A = m$，$\phi = \mathrm{diag}(\phi_1^2, \cdots, \phi_p^2)$，则 A 和 ϕ 的极大似然估计满足方程组：

$$\begin{cases} \mathrm{diag}S = \mathrm{diag}(AA' + \phi) & (2.12) \\ A = S(AA' + \phi)^{-1}A & (2.13) \end{cases}$$

其中

$$S = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})', \quad \bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i.$$

*证明　由于 $\Sigma = AA' + \phi$，此时似然函数为

$$L(\boldsymbol{\mu}, A, \phi) = (2\pi)^{-\frac{np}{2}}|AA' + \phi|^{-\frac{n}{2}}.$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}[(AA' + \phi)^{-1}(nS + n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})(\bar{\boldsymbol{x}} - \boldsymbol{\mu})')]\right\}$$

从而

$$\ln L(\boldsymbol{\mu}, A, \phi) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|AA' + \phi|$$

$$- \frac{1}{2}\mathrm{tr}\{(AA' + \phi)^{-1}[nS$$

$$+ n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})(\bar{\boldsymbol{x}} - \boldsymbol{\mu})']\}$$

利用第一章第四节中有关矩阵的微分公式：

(i) $d\ln|X| = \mathrm{tr}X^{-1}dX$

(ii) $dX^{-1} = -X^{-1} \cdot dX \cdot X^{-1}$

(iii) $df(X) = \mathrm{tr}\left(\frac{\partial f}{\partial X}\right)' dX$

(iv) $d(\boldsymbol{x}'A\boldsymbol{x}) = 2(A\boldsymbol{x})'d\boldsymbol{x}$

(v) $\mathrm{tr}[Ad(XX')] = \mathrm{tr}[2(AX)'dX]$

上述五个公式中 A 表示对称矩阵，$\boldsymbol{x}$ 表示向量，X 表示矩阵。此外，还注意到 $\mathrm{tr}(AB) = \mathrm{tr}(BA) = \mathrm{tr}(AB)'$，从而有

$$d\ln L(\boldsymbol{\mu},A,\phi) = -\frac{n}{2}\operatorname{tr}(AA'+\phi)^{-1}(dAA'+d\phi)$$

$$+\operatorname{tr} n(\bar{x}-\boldsymbol{\mu})'(AA'+\phi)^{-1}d\boldsymbol{\mu}$$

$$+\frac{1}{2}\operatorname{tr}\{(AA'+\phi)^{-1}(dAA'+d\phi)(AA'+\phi)^{-1}$$

$$\cdot[nS+n(\bar{x}-\boldsymbol{\mu})(\bar{x}-\boldsymbol{\mu})']\}$$

为了简洁和方便，记 $f=\ln L(\boldsymbol{\mu},A,\phi),\Sigma=AA'+\phi$，$W=nS+n(\bar{x}-\boldsymbol{\mu})(\bar{x}-\boldsymbol{\mu})'$，经整理后得

$$df = n\operatorname{tr}(\bar{x}-\boldsymbol{\mu})'\Sigma^{-1}d\boldsymbol{\mu}-\operatorname{tr}[n\Sigma^{-1}A-\Sigma^{-1}W\Sigma^{-1}A]'dA$$

$$-\frac{1}{2}\operatorname{tr}[n\Sigma^{-1}-\Sigma^{-1}W\Sigma^{-1}]d\phi$$

令 f 关于 $\boldsymbol{\mu},\phi,A$ 的导数为 0，再注意到 $d\phi=\operatorname{diag}(d\phi_1^2,\cdots,d\phi_p^2)$，有

(i) $\hat{\boldsymbol{\mu}}=\bar{x}$

(ii) $$\operatorname{diag}[(AA'+\phi)^{-1}-(AA'+\phi)^{-1}S(AA'+\phi)^{-1}]=0 \tag{2.14}$$

(iii) $$(AA'+\phi)^{-1}A-(AA'+\phi)^{-1}S(AA'+\phi)^{-1}A=0 \tag{2.15}$$

由(2.15)得到

$$(AA'+\phi)^{-1}[I-S(AA'+\phi)^{-1}]A=0$$

从而有

$$A=S(AA'+\phi)^{-1}A$$

这就得到了(2.13)式。为了化简(2.14)式，注意到(2.14)等价于

$$\operatorname{diag}\{\phi(AA'+\phi)^{-1}[I-S(AA'+\phi)^{-1}]\phi\}=0 \tag{2.16}$$

由(2.13)式知$AA'=S(AA'+\phi)^{-1}AA'$，从而$\phi=S-AA'$，所以有

$$S(AA'+\phi)^{-1}\phi=S-AA'=\phi(AA'+\phi)^{-1}S \tag{2.17}$$

由于

$$\phi(AA'+\phi)^{-1}[I-S(AA'+\phi)^{-1}]\phi$$

$$=\phi(AA'+\phi)^{-1}\phi-\phi(AA'+\phi)^{-1}(S-AA')$$

$$= \phi(AA'+\phi)^{-1}\phi - (S - AA') + \phi(AA'+\phi)^{-1}AA'$$
$$= \phi + AA' - S$$

于是得到

$$\mathrm{diag}S = \mathrm{diag}(\phi + AA')$$

这便是(2.12)式，从而定理得证。∎

由上一段的讨论知道载荷矩阵 A 是不唯一，为此有人建议[26]添加一个计算上方便的唯一性条件:

$$A'\phi^{-1}A = \Delta \tag{2.18}$$

上式中的 Δ 是一个对角阵。

方程组(2.12),(2.13)，再加上唯一性条件(2.18)一般只能用迭代法来求解。但这种迭代法是否收敛并不知道。 直到1967年 Jöreskog 等人提出一个较为实用的计算方法，极大似然法才被人们所采用。下面我们简单的介绍 Jöreskog 提出的方法。

定理 2.2 设 $x_1, \cdots, x_n$ 为正态总体 $N_p(\mu, \Sigma)$ $(\Sigma > 0)$ 的样本 $(n > p)$、Σ 具有结构: $\Sigma = AA' + \phi$，且 A 和 ϕ 满足唯一性条件(2.18)，则 A 与 ϕ 的极大似然估计满足

$$(\phi^{-\frac{1}{2}}S\phi^{-\frac{1}{2}})(\phi^{-\frac{1}{2}}A) = (\phi^{-\frac{1}{2}}A)(I + \Delta) \tag{2.19}$$

$$\phi = \mathrm{diag}(S - AA') \tag{2.20}$$

且

$$\mathrm{tr}[(AA' + \phi)S^{-1}] = p \tag{2.21}$$

上述定理的证明类似于定理 2.1，在[26]给出了详细的证明。此定理及其证明中，读者要注意到:

(1) (2.19)式表明 $\phi^{-\frac{1}{2}}A$ 的第 i 列是矩阵 $\phi^{-\frac{1}{2}}S\phi^{-\frac{1}{2}}$ 对应于特征根 $1 + \Delta_i$ 的特征向量(不是单位化)，其中

$$S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})',$$

且 $\Delta_1 \geqslant \Delta_2 \geqslant \cdots \geqslant \Delta_m$.$\Delta_i$ 为 Δ 的对角元素。

(2) 在定理的证明中，先把极大似然估计 $\hat{\mu} = \bar{x}$ 代入似然函数，并对似然函数取对数，略去与 A 和 ϕ 无关的常数项，问题归结

为，在唯一性条件：$A'\phi^{-1}A=\Delta$（为对角阵）之下，求

$$L^*(\hat{\mu},A,\phi)=\ln|\Sigma|-\ln|S|+\operatorname{tr}(\Sigma^{-1}S)-p \quad (2.22)$$

的最小值。

迭代步骤为：

(i) 计算特殊方差的初始估计。Jöreskog 建议取

$$\hat{\phi}_i=\left(1-\frac{1}{2}\cdot\frac{m}{p}\right)\cdot\left(\frac{1}{s^{ii}}\right) \quad (2.23)$$

其中 s^{ii} 为 S^{-1} 的第 i 个对角元素

(ii) 对于给定的 $\hat{\phi}=\operatorname{diag}(\hat{\phi}_1,\cdots,\hat{\phi}_p)$ 计算

$$S^*=\hat{\phi}^{-\frac{1}{2}}S\hat{\phi}^{-\frac{1}{2}} \quad (2.24)$$

的前 m 个特征根 $\hat{\lambda}_1>\hat{\lambda}_2>\cdots>\hat{\lambda}_m>1$，及相对应的单位化特征向量 $\hat{e}_1,\cdots,\hat{e}_m$，记 $\hat{D}_\lambda=\operatorname{diag}(\hat{\lambda}_1,\cdots,\hat{\lambda}_m)$，$\hat{E}=[\hat{e}_1,\cdots,\hat{e}_m]$，由(2.19)有 $\hat{D}_\lambda=I+\hat{\Delta}$，$\hat{E}=\hat{\phi}^{-\frac{1}{2}}\hat{A}\hat{\Delta}^{-\frac{1}{2}}$。从而得到 A 的估计

$$\hat{A}=\hat{\phi}^{\frac{1}{2}}\hat{E}\hat{\Delta}^{\frac{1}{2}}=\hat{\phi}^{\frac{1}{2}}\hat{E}(\hat{D}_\lambda-I)^{\frac{1}{2}} \quad (2.25)$$

(iii) 将 (ii) 中得到的 $\hat{A}$ 代入目标函数 (2.22) 式，关于 $\phi_1,\cdots,\phi_p$ 使其极小化，得到特殊方差的新估计，将此新估计代入(2.24)，重复步骤 (ii) 和 (iii)，直到相邻两次迭代产生的 $\hat{\phi}$ 和 $\hat{A}$ 无明显差异为止。

2° 主成分法

前面已谈到，我们建立的因子模型使其满足方差结构：$\Sigma=AA'+\phi$。设 Σ 的特征根为 $\lambda_1\geqslant\lambda_2\geqslant\cdots,\geqslant\lambda_p$，相应的特征向量组成的特征向量矩阵为 $P=[e_1,\cdots,e_p]$，记 $D_\lambda=\operatorname{diag}(\lambda_1,\cdots,\lambda_p)$，此时有

$$\Sigma=PD_\lambda P'=\lambda_1e_1e_1'+\lambda_2e_2e_2'+\cdots+\lambda_pe_pe_p'$$

$$=[\sqrt{\lambda_1}\,e_1\,\sqrt{\lambda_2}\,e_2,\cdots,\sqrt{\lambda_p}\,e_p]\begin{bmatrix}\sqrt{\lambda_1}\,e_1'\\ \sqrt{\lambda_2}\,e_2'\\ \vdots\\ \sqrt{\lambda_p}\,e_p'\end{bmatrix}=AA' \quad (2.26)$$

(2.26) 式的分解正是公因子与变量个数一样，特殊因子的方差为 0 的因子模型的方差结构形式，即

$$\Sigma = AA' + O = AA' \tag{2.27}$$

因子分析就是要寻找少数几个公因子来解释变量的相关结构，因此(2.27)的结构形式在实际应用上是无价值的. 类似于主成分的思想，如果 Σ 的最后 $p-m$ 个特征根很小时，在(2.26)式将 $\lambda_{m+1}\boldsymbol{e}_{m+1}\boldsymbol{e}'_{m+1}+\cdots+\lambda_p\boldsymbol{e}_p\boldsymbol{e}'_p$ 略去，这样我们就得到

$$\Sigma \approx [\sqrt{\lambda_1}\boldsymbol{e}_1, \sqrt{\lambda_2}\boldsymbol{e}_2, \cdots, \sqrt{\lambda_m}\boldsymbol{e}_m] \begin{bmatrix} \sqrt{\lambda_1}\boldsymbol{e}'_1 \\ \sqrt{\lambda_2}\boldsymbol{e}'_2 \\ \vdots \\ \sqrt{\lambda_m}\boldsymbol{e}'_m \end{bmatrix} = \hat{A}\hat{A}' \tag{2.28}$$

这里的 $\hat{A}$ 是 $p\times m$ 阶. (2.28)式近似表明了因子模型，(2.1)中的特殊因子是不重要的，能从 Σ 的分解中忽略掉，如果要考虑特殊因子，那么我们可以用 $\mathrm{diag}(\Sigma-\hat{A}\hat{A}')$ 来估计 $\phi=\mathrm{diag}(\phi_1^2,\cdots,\phi_p^2)$，这里的 $\hat{A}$ 是由(2.28)所定义. 此时近似关系为

$$\begin{aligned} \Sigma &\approx \hat{A}\hat{A}' + \mathrm{diag}(\Sigma - \hat{A}\hat{A}') \\ &= [\sqrt{\lambda_1}\boldsymbol{e}_1, \cdots, \sqrt{\lambda_m}\boldsymbol{e}_m] \begin{bmatrix} \sqrt{\lambda_1}\boldsymbol{e}'_1 \\ \vdots \\ \sqrt{\lambda_m}\boldsymbol{e}'_m \end{bmatrix} + \phi \end{aligned} \tag{2.29}$$

其中 $\phi=\mathrm{diag}(\phi_1,\cdots,\phi_p)$，$\phi_i=\sigma_{ii}-\sum_{j=1}^{m}a_{ij}^2\ (i=1,2,\cdots,p)$. 因此载荷矩阵 A 的估计为 $\hat{A}=(\sqrt{\lambda_1}\boldsymbol{e}_1,\cdots,\sqrt{\lambda_m}\boldsymbol{e}_m)$.

在实际问题中 Σ 是未知的，必须用它的样本协方差来代替，但因诸变量的量纲不同，往往将数据标准化，利用标准化数据计算出来的样本协方差便是原数据的相关系数矩阵 R，这样我们将上述

的方法用于 R，可得到类似于(2.29)的表示：

$$R \approx \tilde{A}\tilde{A}' + \tilde{\Phi} = (\sqrt{\tilde{\lambda}_1}\tilde{e}_1, \cdots, \sqrt{\tilde{\lambda}_m}\tilde{e}_m)\begin{bmatrix}\sqrt{\tilde{\lambda}_1}\tilde{e}_1' \\ \vdots \\ \sqrt{\tilde{\lambda}_m}\tilde{e}_m'\end{bmatrix} + \begin{pmatrix}\tilde{\phi}_1^2 & & 0 \\ & \ddots & \\ 0 & & \tilde{\phi}_p^2\end{pmatrix} \tag{2.30}$$

其中 $\tilde{\lambda}_1 \geqslant \tilde{\lambda}_2 \geqslant \cdots \geqslant \tilde{\lambda}_m, \tilde{e}_1, \cdots, \tilde{e}_m$. 分别是 R 的前 m 个较大特征根与相应的单位化正交特征向量，特殊方差 $\tilde{\Phi}$ 用 $\mathrm{diag}(R-\tilde{A}\tilde{A}')$ 来估计。因此，由 R 出发因子分析模型的载荷矩阵的估计为

$$\tilde{A} = (\sqrt{\tilde{\lambda}_1}\tilde{e}_1, \cdots, \sqrt{\tilde{\lambda}_m}\tilde{e}_m) \tag{2.31}$$

特殊因子的方差 $\phi_i^2 (i = 1, 2, \cdots, p)$ 的估计为

$$\tilde{\phi}_i^2 = 1 - \sum_{j=1}^{m} \tilde{a}_{ij}^2 \tag{2.32}$$

其中 $\tilde{a}_{ij}$ 为 $\tilde{A}$ 的 (i,j) 元素。

共同度 h_i 的估计为

$$\tilde{h}_i^2 = \sum_{j=1}^{m} \tilde{a}_{ij}^2 \quad (i = 1, 2, \cdots, p) \tag{2.33}$$

公因子 F_j 对各变量的“贡献” g_j^2 的估计为

$$\tilde{g}_j^2 = \sum_{i=1}^{p} \tilde{a}_{ij}^2, \ (j = 1, 2, \cdots, m) \tag{2.34}$$

公因子数 m 的确定仍按主成分的思想来定，比较理想的情况是只有少数几个公因子对变量的“贡献”很大，注意到在标准化变量的情况下，所有变量的总方差为 $\mathrm{tr}(R) = p$，第 j 个公因子 F_j 的“贡献”为

$$\tilde{g}_j^2 = \sum_{i=1}^{m} \tilde{a}_{ij}^2 = (\sqrt{\tilde{\lambda}_j}\tilde{e}_j')(\sqrt{\tilde{\lambda}_j}\tilde{e}_j) = \tilde{\lambda}_j \tag{2.35}$$

因而寻找一个 m 使得

$$\left(\sum_{i=1}^{m} \lambda_i \Big/ p\right) 100\% \geqslant 80\% \text{（或 }75\%\text{）} \tag{2.36}$$

就确定该 m 为公因子数.

在上一节中我们提到 Q 型的因子模型(1.7),它是将样品表示成公因子和特殊因子的线性组合，通过少数几个公因子来研究样品相似性的内在联系. Imbrie [21]提出了用样品的相似系数矩阵进行 Q 型因子分析. 样品的相似系数矩阵为

$$Q = (\cos\theta_{ik})_{n\times n} \tag{2.37}$$

其中

$$\cos\theta_{ik} = \frac{\sum_{t=1}^{p} x_{ti}\cdot x_{tk}}{\sqrt{\sum_{t=1}^{p} x_{ti}^2 \cdot \sum_{t=1}^{p} x_{tk}^2}} \tag{2.38}$$

表示样品 x_i 与 x_k 的夹角余弦。

类似于 R 型分析,求 Q 的特征根 $\tilde{\lambda}_1 \geqslant \tilde{\lambda}_2 \geqslant \cdots \geqslant \tilde{\lambda}_n$ 及相应的单位化正交特征向量 $\tilde{e}_1, \tilde{e}_2, \cdots, \tilde{e}_n$, 选择 m 使得

$$\left(\sum_{i=1}^{m} \tilde{\lambda}_i \Big/ \sum_{i=1}^{n} \tilde{\lambda}_i\right) 100\% \geqslant 80\% \text{（或 }75\%\text{）},$$

这个 m 做为公因子的数目，这样便可得到 Q 型因子分析的载荷矩阵的估计为

$$\tilde{A} = (\sqrt{\tilde{\lambda}_1}\,\tilde{e}_1, \cdots, \sqrt{\tilde{\lambda}_m}\,\tilde{e}_m) \tag{2.39}$$

无论是 R 型或 Q 型因子分析,我们要对公因子给予具体解释时,都必须通过载荷阵找出各公因子的“代表”变量或“代表”样品，通过“代表”变量或“代表”样品的特征给予公因子具体解释或赋予公因子具体含义。

（三） 协方差矩阵结构的检验

我们建立的因子分析模型，使其满足协方差阵具有结构 $\Sigma = AA' + \Phi$,这里 A 是 $p\times m$ 阶矩阵. p 个变量的协方差阵是否存

在这样的结构呢？或者说，因子模型是否适用？这就需要检验假设。

$$H_0: \Sigma = AA' + \phi; \ H_1: \Sigma \text{ 不取 } AA' + \phi \text{ 的形式}. \tag{2.40}$$

我们就正态总体来讨论此问题。

定理 2.3 设 $x_1, \cdots, x_n$ 是正态总体 $N_p(\mu, \Sigma), (\Sigma > 0)$ 的样本 $(n > p)$。检验假设(2.40)的检验似然比统计量为

$$\lambda^* = \frac{|S|^{n/2}}{|\hat{A}\hat{A}' + \hat{\phi}|^{n/2}} \tag{2.41}$$

其中 $S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'$，$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$，$\hat{A}$ 和 $\hat{\phi}$ 是 A 和 ϕ 的极大似然估计，即 $\hat{A}$ 和 $\hat{\phi}$ 是方程(2.12)，(2.13)的解。

此定理的证明是利用大家熟知的似然比方法便可完成。(2.41)的等价统计量为

$$\lambda = |S| / |\hat{A}\hat{A}' + \hat{\phi}| \tag{2.42}$$

有关 λ 的渐近分布是有如下结论。

定理 2.4 当 $n \to \infty$ 时

$$-\left[n - \frac{2p + 11}{6} - \frac{2}{3}m\right] \ln \lambda \sim \chi^2(f) \tag{2.43}$$

其中 $f = \frac{1}{2}[(p - m)^2 - (p + m)]$.

应用上面的检验时，必须先求出极大似然估计 $\hat{A}$ 和 $\hat{\phi}$，这须要用计算机有关程序进行迭代计算，在进行计算时仍然涉及到事先要取多少个公因子，即 m 应取多大的问题。有的人是经验地定一个 m，有的人是利用主成分方法计算相关矩阵的特征根，由前 m 个特征根在所有特征根中所占的比例来确定 m，有的人把大于1的特征根的个数定为 m。一旦 m 确定了，就计算 $\hat{A}$ 和 $\hat{\phi}$，然后进行检验假设(2.40)，如果接受 H_0，说明模型适用，如果拒绝 H_0，那么再增加一个公因子，即 $m + 1$ 个公因子，再作以上的估计和检

验，由于$m<p$的，总可以找到一个适当的公因子数目，使得相应的因子模型适用。当然，在许多应用实例中，人们往往直接利用主成分法定出 m，而不再去作检验。

（四） 实例

下面的两个实例都是利用主成分估计方法进行因子分析。

例 2.1 在某市高考的理科考生中，抽取了 90 名考生的成绩作R型因子分析，变量为七门考试科目，利用计算机得到这七门科目的相关系数矩阵(表 2.1)，计算相关系数矩阵的特征根及相应的单位化特征向量列在表 2.2。

表 2.1 七门考试科目的相关系数矩阵

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1：政治	1						
x_2：语文	0.35	1					
x_3：数学	0.15	0.34	1				
x_4：物理	0.29	0.39	0.49	1			
x_5：化学	0.43	0.34	0.55	0.54	1		
x_6：外语	0.40	0.43	0.25	0.34	0.35	1	
x_7：生物	0.49	0.56	0.42	0.49	0.49	0.2	1

表 2.2 相关系数矩阵的特征根及单位化特征向量

特征根 λ_i	3.51	1.03	0.72	0.56	0.50	0.36	0.32
累积百分比	50.[illegible]%	64.9%	75.1%	83.1%	90.3%	95.4%	100%
	e_1	e_2	e_3	e_4	e_5	e_6	e_7
特征向量	0.33	−0.44	−0.66	−0.30	0.04	0.32	−0.27
	0.37	−0.20	0.54	−0.65	0.00	−0.29	−0.17
	0.34	0.59	0.13	0.00	0.52	0.43	−0.27
	0.38	0.35	0.00	0.08	−0.83	0.14	−0.13
	0.40	0.31	−0.43	0.04	0.17	−0.70	0.19
	0.37	−0.41	0.23	0.69	0.09	−0.15	−0.38
	0.45	−0.20	0.14	0.09	0.06	0.31	0.79

从表 2.2 可以看出前三个特征根的累积百分比为 75.1%，所以提取三个公因子，利用(2.31)～(2.34)可分别计算出因子载荷阵，特殊因子方差、共同度及公因子方差贡献．并列于表 2.3.

表 2.3　公因子载荷阵、共同度、特殊因子方差及公因子方差贡献

变量 \ 载荷 \ 公因子	F_1	F_2	F_3	共同度	特殊因子方差
x_1：政治	0.62	−0.44	−0.56	0.88	0.12
x_2：语文	0.69	−0.20	0.46	0.73	0.27
x_3：数学	0.64	0.60	0.11	0.78	0.22
x_4：物理	0.71	0.35	0.00	0.62	0.38
x_5：化学	0.75	0.31	−0.37	0.71	0.29
x_6：外语	0.69	−0.41	0.20	0.69	0.31
x_7：生物	0.84	−0.20	0.12	0.75	0.25
公因子贡献	3.51	1.03	0.72		
贡献累积	50.1%	64.9%	75.1%		

由于载荷 a_{ij} 表示了第 i 个变量与第 j 个公因子的相关系数，因此我们可以根据各个公因子上的载荷对公因子赋予解释．先看公因子 F_1，所有七个变量在它上面都有相差不太大的正载荷，这个公因子可赋予它为中学生的“一般智力”因子，公因子 F_2 与 F_3 其载荷有正有负，暂不好解释．留到下一节的正交旋转后再来作较为详细的解释．

例 2.2　食品部门为了掌握男、女各年龄组的人对各种食品的喜爱情况，调查了男、女十个不同年龄组的人对 100 种食物(包括主食、面条、面包、肉类、鱼类、蛋类及各种蔬菜等)的喜欢食用程度的数据，数据采用打分形式，10 分为满分，然后对其进行因子分析．变量为 x_1＝男 16 岁以下，x_2＝男 16～20 岁，x_3＝男 21～30 岁，x_4＝男 31～40 岁，x_5＝男 41 岁以上，x_6＝女 16 岁以下，x_7＝女 16～20 岁，x_8＝女 21～30 岁，x_9＝女 31～40 岁，x_{10}＝女 41 岁以上．每个年龄组调查了 50～120 人不等，然后再

表 2.4　十个年龄组食品嗜好的相关系数矩阵

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	1									
x_2	0.8708	1								
x_3	0.5158	0.7588	1							
x_4	0.3701	0.6042	0.8524	1						
x_5	0.1722	0.4021	0.7262	0.8742	1					
x_6	0.9384	0.8207	0.5165	0.3580	0.2077	1				
x_7	0.8107	0.8381	0.6584	0.4875	0.3543	0.8888	1			
x_8	0.6106	0.7096	0.6990	0.6199	0.5235	0.7465	0.8949	1		
x_9	0.5004	0.6470	0.7013	0.7207	0.7101	0.6215	0.7678	0.8528	1	
x_{10}	0.3298	0.4570	0.5584	0.6321	0.7479	0.4932	0.6415	0.7741	0.9112	1

表2.5 不同年龄组的因子载荷、特殊因子方差、共同度及公因子方差贡献

变量 \ 载荷 \ 公因子	F_1	F_2	F_3	共同度	特殊因子方差
x_1	0.7423	−0.5743	−0.1538	0.9045	0.0955
x_2	0.8604	−0.3099	−0.2742	0.9115	0.0885
x_3	0.8318	0.2064	−0.4354	0.9241	0.0759
x_4	0.7799	0.4706	−0.3294	0.9382	0.0618
x_5	0.6740	0.6455	−0.1093	0.8829	0.1171
x_6	0.8050	−0.5340	0.0728	0.9385	0.0615
x_7	0.8970	−0.3288	0.1423	0.9330	0.0670
x_8	0.8947	−0.0379	0.2090	0.8504	0.1496
x_9	0.8993	0.2168	0.2604	0.9236	0.0764
x_{10}	0.7926	0.3571	0.4441	0.9530	0.0470
公因子贡献	6.828	1.762	0.754		
累积百分比	68.28%	85.90%	93.44%		

取平均。从所调查的数据，计算得十个年龄组的相关系数矩阵如表 2.4 所示。从相关系数矩阵中可看出同年龄组的男女之间相关系数比较大，特别是男女少年 (16 岁以下)相关系数高达 0.9383。

计算相关矩阵的特征根为：6.828，1.762，0.754，0.262，0.122，0.098，0.072，0.044，0.036，0.022。头三个特征根的累积百分比已达 93.44%，因此我们取三个公因子，前三个特征根对应的单位化特征向量为

$$\hat{e}_1 = (0.2841, 0.3293, 0.3183, 0.2985, 0.2579, 0.3081, 0.3433, 0.3424, 0.3442, 0.3033)'$$

$$\hat{e}_2 = (-0.4357, -0.2351, 0.1571, 0.3582, 0.4913, -0.4065, -0.2503, -0.0288, 0.1650, 0.2718)'$$

$$\hat{e}_3 = (-0.1771, -0.3158, -0.5014, -0.3794, -0.1259, 0.0838, 0.1639, 0.2407, 0.2999, 0.5115)'$$

根据(2.31)～(2.34)可分别计算因子载荷阵、特殊因子方差、共同

度及各公因子方差贡献，并列于表 2.5。

公因子 F_1，男、女十个年龄组在它上面的载荷都为正且在 0.8 左右，所反映的是不分性别和年龄的共同嗜好程度；公因子 F_2，在同一性别的年龄组之间载荷有正负之别，它反映同一性别不同年龄组之间存在不同嗜好；公因子 F_3，男性载荷为负，女性载荷为正，反映了男与女之间的不同嗜好。对于这三个公因子更精确的解释将留到正交旋转以后再来讨论。

（五） 因子模型的几何解释

为了便于理解，我们对因子模型作一点几何解释。

在因子分析中，我们可以把互不相关的，即两两之间的相关系数(夹角余弦)为零，各自方差为 1 的 m 个公因子和 p 个特殊因子当成 $m+p$ 个相互垂直的单位向量，以它们为坐标轴构成 $m+p$ 维空间的一个直角坐标系．并称为因子空间．于是变量 x_i 可以用向量 $p_i=(a_{i1},a_{i2},\cdots,a_{im},0,\cdots 0,\phi_i,0,\cdots 0)'$ 表示，其中 ϕ_i 是 x_i 在对应于自己的特殊因子轴上的载荷．由于 x_i 标准化，显然 p_i 的长度为

$$\|p_i\|=(a_{i1}^2+a_{i2}^2+\cdots+a_{im}^2+\phi_i^2)^{\frac{1}{2}}=\sqrt{\mathrm{var}\, x_i}=1$$

此时 p_i 与各个因子轴 F_j 的夹角余弦为

$$\cos(\widehat{p_iF_j})=\|p_i\|\cos(\widehat{p_iF_j})=a_{ij}=r_{p_i,F_j}$$

这表明了 p_i 与各公因子的夹角余弦就等于其对应的坐标，也就是等于变量 x_i 与各公因子的相关系数。此外，对于 x_i 和 x_j 所对应的向量 p_i 与 p_j 的夹角余弦为

$$\cos(\widehat{p_ip_j})=\frac{p_i'p_j}{\|p_i\|\|p_j\|}=p_i'p_j=\sum_{t=1}^{m}a_{it}a_{jt}=r_{x_i,x_j}$$

它恰好等于变量 x_i 与 x_j 的相关系数。

§ 10.3 因子正交旋转

（一） 方差极大旋转

在上节(一)段中已经指出满足方差结构 $\Sigma=AA'+\phi$ 的因

子模型中的公因子与载荷矩阵并不唯一。若A满足 $\Sigma = AA' + \psi$，那么对于$m \times m$阶正交矩阵 T，则 $F^* = T'F$ 也是公因子，相应的载荷矩阵为 $A^* = AT$，它们也满足上述方差结构形式，而 $A^* = AT$ 表明新的载荷阵 A^* 是将A经过一个正交旋转而得到，而这种旋转对共同度是不变的，从而对特殊因子方差也不变。为了方便叙述我们把A称为初始载荷矩阵。在实际应用中要对公因子 F_j 进行解释，因此希望各个变量在该因子上的载荷 $a_{ij}(i = 1, 2, \cdots, p)$ 两极分化，即只有少数变量在 F_j 上有较大的载荷，而其余变量的载荷尽可能的小，这样 F_j 的具体含义可由载荷较大的变量根据具体问题加以解释。若载荷矩阵的每一列都具有两极分化的特征，称这种载荷阵具有简单结构。由以上所述，我们要通过正交旋转寻找一个简单结构的载荷矩阵。在这里我们介绍一种方差极大旋转。

方差极大旋转是使载荷阵的每一列元素平方的方差达到最大作为载荷阵简单结构的准则。对于公因子 $F_j(j = 1, 2, \cdots, m)$，p 个变量在它上面的载荷值平方的方差为

$$V_j = \frac{1}{p}\sum_{i=1}^{p}(b_{ij}^2 - \bar{b}_j^2)^2 = \left[p\sum_{i=1}^{p}(b_{ij}^2)^2 - \left(\sum_{i=1}^{p}b_{ij}^2\right)^2\right]\Big/p^2$$

其中 b_{ij} 表示经过正交旋转后所得因子载荷矩阵B的元素，$\bar{b}_j^2 = \frac{1}{p}\sum_{i=1}^{p}b_{ij}^2$。对于所有$m$个公因子载荷值平方的总方差为

$$V^* = \sum_{j=1}^{m}V_j = \sum_{j=1}^{m}\left[p\sum_{i=1}^{p}(b_{ij}^2)^2 - \left(\sum_{i=1}^{p}b_{ij}^2\right)^2\right]\Big/p^2 \quad (3.1)$$

旋转后的载荷矩阵的元素就是要使得(3.1)式达到最大. 考虑到各变量共同度 $h_i^2(i = 1, 2, \cdots, p)$ 的差异的不平衡，我们用 b_{ij}^2/h_i^2 代替(3.1)式中的 b_{ij}^2，从而得到

$$V = \sum_{j=1}^{m}\left[p\sum_{i=1}^{p}(b_{ij}^2/h_i^2)^2 - \left(\sum_{i=1}^{p}(b_{ij}^2/h_i^2)\right)^2\right]\Big/p^2 \quad (3.2)$$

这样旋转的准则就是使得 (3.2) 达到最大。现在的问题就归结为寻找一个 $m\times m$ 阶的正交矩阵 T 使得 $B=AT$ 且使(3.2)达到最大。m 维坐标系的正交旋转可以由所有的两两公因子轴构成的平面正交旋转一个角度来完成，一共有 C_m^2 个因子轴组成的平面，每次的转角 φ 都必须使 V 达到最大。

对任意的两个因子轴 F_g 及 $F_q(g=1,2,\cdots,m-1.\ g<q)$ 所构成的因子平面正交旋转一个角度 $\varphi_{g,q}$，变量 $x_i(i=1,2,\cdots,p)$ 在 F_g 和 F_q 轴上的初始载荷为 (a_{ig},a_{iq})，经旋转后的载荷为 (b_{ig},b_{iq})，$(i=1,2,\cdots,p)$，两者之间的关系为

$$\begin{pmatrix} b_{1g} & b_{1q} \\ b_{2g} & b_{2q} \\ \vdots & \vdots \\ b_{pg} & b_{pq} \end{pmatrix}=\begin{pmatrix} a_{1g} & a_{1q} \\ a_{2g} & a_{2q} \\ \vdots & \vdots \\ a_{pg} & a_{pq} \end{pmatrix}\begin{pmatrix} \cos\varphi_{gq} & -\sin\varphi_{gq} \\ \sin\varphi_{gq} & \cos\varphi_{gq} \end{pmatrix}$$

上面只表明旋转 F_g 与 F_q 轴得到 $x_i(i=1,2,\cdots,p)$ 在两个新公因子轴上的载荷 b_{ig} 和 b_{iq}。此时 $x_i,(i=1,2,\cdots,p)$ 在其它公因子轴上的载荷并没有改变，因此，当 F_g 与 F_q 轴旋转后得到的载荷矩阵应写成

$$\widetilde{B}=A\cdot T_{g\cdot q} \tag{3.3}$$

其中 $\widetilde{B}$ 为将 A 中的 g，q 两列元素改为 $(b_{1g},\cdots,b_{pg})'$ 和 $(b_{1q},\cdots,b_{pq})'$，其余元素不变，即

$$\left.\begin{aligned} & b_{ig}=a_{ig}\cos\varphi_{gq}+a_{iq}\sin\varphi_{gq} \\ & b_{iq}=-a_{ig}\sin\varphi_{gq}+a_{iq}\cos\varphi_{gq} \\ & b_{it}=a_{it}\quad t\neq g,q, \\ & i=1,2,\cdots,p. \end{aligned}\right\} \tag{3.4}$$

而 $T_{g,q}$ 为 $m\times m$ 阶矩阵，它的第 g 行第 g 列的元素为 $\cos\varphi_{gq}$，第 g 行第 q 列的元素为 $-\sin\varphi_{g,q}$，第 q 行第 g 列的元素为 $\sin\varphi_{g,q}$，第 q 行第 q 列的元素为 $\cos\varphi_{gq}$，除此之外的其它元素——对角线为 1，非对角线为 0。

我们必须对所有的公因子轴配对旋转，总共要旋转 C_m^2 次，即

得到

$$B_{(1)}=AT_{12}\cdot T_{13}\cdots T_{1m}\cdot T_{23}\cdots T_{2m}\cdots T_{(m-1)m}$$

$$=A\prod_{g=1}^{m-1}\prod_{q=g+1}^{m}T_{gq}=AC_1 \tag{3.5}$$

其中 $C_1=\prod_{g=1}^{m-1}\prod_{q=g+1}^{m}T_{gq}$。由于正交阵的乘积仍为正交阵，故 C_1 为一个 $m\times m$ 阶正交阵。$B_{(1)}$ 为第一个循环完成后的载荷矩阵。

例如，设有三个公因子 F_1,F_2,F_3。初始载荷矩阵为

$$A=\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ \vdots & \vdots & \vdots\\ a_{p1} & a_{p2} & a_{p3}\end{bmatrix}$$

对 F_1—F_2 进行旋转，T_{12} 为

$$T_{12}=\begin{bmatrix}\cos\varphi_{12} & -\sin\varphi_{12} & 0\\ \sin\varphi_{12} & \cos\varphi_{12} & 0\\ 0 & 0 & 1\end{bmatrix}$$

对 F_1—F_3 进行旋转，T_{13} 为

$$T_{13}=\begin{bmatrix}\cos\varphi_{13} & 0 & -\sin\varphi_{13}\\ 0 & 1 & 0\\ \sin\varphi_{13} & 0 & \cos\varphi_{13}\end{bmatrix}$$

对 F_2—F_3 进行旋转，T_{23} 为

$$T_{23}=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\varphi_{23} & -\sin\varphi_{23}\\ 0 & \sin\varphi_{23} & \cos\varphi_{23}\end{bmatrix}$$

此时 $B_{(1)}=AT_{12}T_{13}T_{23}=AC_1$。

将第一个循环完成后的载荷矩阵 $B_{(1)}$ 的元素代入(3.2)可计算得 $V_{(1)}$。

从 $B_{(1)}$ 出发重复上述的步骤进行第二个循环的旋转得到

$$B_{(2)}=B_{(1)}\prod_{g=1}^{m-1}\prod_{q=g+1}^{m}T_{gq}=B_{(1)}C_2 \tag{3.6}$$

将 $B_{(2)}$ 的元素代入(3.2)计算得 $V_{(2)}$。再从 $B_{(2)}$ 出发进行第三循环…，这样不断地重复，便可得到 V 的一个非降序列 $V_{(1)}<V_{(2)}<\cdots$。要循环多少次呢？对于给定的一个小正数 δ，只要满足 $V_{(k)}-V_{(k-1)}<\delta$ 时就停止，最后就得到

$$B_{(k)}=B_{(k-1)}C_k=B_{(k-2)}C_{k-1}\cdot C_k=\cdots=AC_1\cdots C_k$$

$$=A\prod_{i=1}^{k}C_i=AC \tag{3.7}$$

$B_{(k)}$ 就是我们所要的具有简单结构的载荷阵。通常称它为经过正交旋转的载荷阵。

进行旋转时，旋转角 φ_{gq} 如何确定呢？因为我们的旋转是要使(3.2)达到最大，因此对因子轴 F_g 和 F_q 进行旋转的转角 φ_{gq} 应根据极值原理来确定：

(1) 将(3.4)代入(3.2)，V 就成为 φ_{gq} 的一个函数，并对 φ_{gq} 求一阶导数令其为 0，可解得

$$\operatorname{tg}4\varphi_{gq}=\frac{d-2ab/p}{c-(a^2-b^2)/p}=\frac{\alpha}{\beta} \tag{3.8}$$

其中 p 为变量的个数，$\alpha=d-2ab/p$，$\beta=c-(a^2-b^2)/p$ 而

$$a=\sum_{i=1}^{p}\mu_i,\quad b=\sum_{i=1}^{p}\nu_i,\quad c=\sum_{i=1}^{p}(\mu_i^2-\nu_i^2),\quad d=2\sum_{i=1}^{p}\mu_i\nu_i.$$

这里的 μ_i 和 ν_i 分别为

$$\mu_i=(a_{ig}/h_i)^2-(a_{iq}/h_i)^2,\quad \nu_i=2(a_{ig}/h_i)(a_{iq}/h_i)$$

注意，若是确定某个循环的第 s 次旋转，这里的 a_{ig}/h_i 及 a_{iq}/h_i 应理解为该循环的第 $s-1$ 次旋转后得到的载荷矩阵元素。

(2) 将(3.2)展开，把包含 φ_{gq} 的项合并、化简，最后只剩下包含 $\sin 4\varphi_{gq}$ 和 $(\sin 2\varphi_{gq})^2=\frac{1}{2}-\frac{1}{2}\cos 4\varphi_{gq}$ 的项，显然它

们都是以 $\frac{\pi}{2}$ 为周期的函数，因此(3.8)式的 φ_{gq} 只须在 $\frac{\pi}{2}$ 范围内考虑就行，通常在 $\left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$ 之间考虑。另外，由于要使 V 达到最大，对(3.2)式对 $\varphi_{g,q}$ 求二阶导数，由二阶导数小于零可得到 $\frac{1}{\alpha}\sin 4\varphi_{g,q} > 0$，又因为 $\sin(-4\varphi_{gq}) = -\sin 4\varphi_{gq}$，所以 φ_{gq} 的符号应与 α 同号，由此可按(3.8)式中 β 与 α 的符号来确定 φ_{gq} 的符号，即因子轴应向逆时针方向还是顺时针方向旋转。表 3.1 给出了根据 β 和 α 的符号来确定 φ_{gq} 应在哪一个象限。

表 3.1 旋转角 φ_{gq} 所在的象限

α	β	象　限	$4\varphi_{gq}$ 值范围	φ_{gq} 值的范围	转角度数
+	+	1	$0\sim\pi/2$	$0\sim\pi/8$	$0°\sim22.5°$
+	−	2	$\pi/2\sim\pi$	$\pi/8\sim\pi/4$	$22.5°\sim45°$
−	−	3	$-\pi\sim-\pi/2$	$-\pi/4\sim-\pi/8$	$-45°\sim-22.5°$
−	+	4	$-\pi/2\sim0$	$-\pi/8\sim0$	$-22.5°\sim0$

例如，$\mathrm{tg}\,4\varphi_{gq} = \frac{17.784}{-3.672} = -4.843$，由表 3.1 知 φ_{gq} 应在 $22.5°\sim45°$ 之间，查三角函数表知 $4\varphi_{gq} = 101°40'$，$\varphi_{gq} = 25°25'$。

对于 Q 型因子分析的因子轴旋转和上述的方法完全一样，但要将(3.2)式中的 p 改为样品数目 n 即可。

例 3.1 (续例 2.1)　对某市高考的理科考生中，抽取了 90 名考生的成绩进行了 R 型因子分析，为了便于解释各公因子的含义，我们将例 2.1 中得到的初始载荷矩阵进行正交旋转，其旋转后的载荷矩阵列在表 3.2 中。

经过旋转后，原来的公因子 F_1，即所谓的"一般智力"因子分别渗透到各个因子中去。从表 3.2 的载荷矩阵我们看到：

(1) 在公因子 F_1 中，政治有压倒多数的正载荷 0.87。故 F_1 可解释为"中学生政治科"因子（分析和解释社会现实及本人思想

表 3.2 正交旋转后的载荷矩阵

变量 \ 载荷 \ 公因子	F_1	F_2	F_3	共同度	特殊因子方差
x_1: 政治	0.87	−0.07	0.35	0.88	0.12
x_2: 语文	−0.05	0.18	0.84	0.73	0.27
x_3: 数学	0.03	0.84	0.29	0.78	0.22
x_4: 物理	0.22	0.66	0.38	0.62	0.38
x_5: 化学	0.48	0.65	0.24	0.71	0.29
x_6: 外语	0.24	−0.01	0.80	0.69	0.31
x_7: 生物	0.32	0.25	0.78	0.75	0.25
公因子贡献	1.21	1.67	2.37		
贡献率	17.3%	23.86%	33.86%		

的适应能力)。但要注意到该因子的方差贡献率为 17.3%，在旋转后的公因子中占第三位。

(2) 在公因子 F_2 中，数理化三科在该因子上载荷都比较大，所以 F_2 解释为"数理化能力"因子(抽象思维、逻辑推理、运算和操作能力)。人们常常把数理化看成有紧密联系的科目，这一点从统计分析的角度上看就更加明显。其方差贡献率为 23.86%，为第二位。

(3) 在公因子 F_3 中，语文、外语和生物在该因子上载荷都比较大，该因子可解释为"语外生能力"因子(表达、叙述、想象和记忆能力)，其方差贡献率为 33.86% 占首位，因此语、外、生在中学生教育中必须加强。

由上述的分析得出的知识与能力结构同中学的传统分类基本吻合。这三个公因子的累积贡献率达 75%，它们基本上解释了每个变量的变异的绝大部分，有些变量的特殊方差稍大一点，如物理(0.38)和外语(0.31)，这说明物理、外语与其它科目比较起来需要学生具备一些特殊的素质，比如实验能力、语言习惯等。

例 3.2 (续例 2.2) 食品部门对男、女各年龄组的爱好调查，

表 3.3 正交旋转后不同年龄组的因子载荷

变量 \ 载荷 \ 公因子	F_1	F_2	F_3	共同度	特殊因子方差
x_1	0.9372	-0.0871	0.1363	0.9045	0.0955
x_2	0.8415	-0.1304	0.4318	0.9115	0.0885
x_3	0.4547	-0.2207	0.8177	0.9241	0.0759
x_4	0.2268	-0.2781	0.8997	0.9382	0.0618
x_5	-0.0056	-0.4527	0.8234	0.8829	0.1171
x_6	0.9143	-0.3134	0.0645	0.9385	0.0615
x_7	0.8150	-0.4781	0.2003	0.9330	0.0670
x_8	0.5908	-0.6176	0.3395	0.8504	0.1496
x_9	0.4179	-0.7273	0.4690	0.9236	0.0764
x_{10}	0.1981	-0.8719	0.3918	0.9530	0.0470
贡献率	38.9%	23.5%	28.9%		

在例 2.2 中进行 R 型因子分析，为了更精确解释各个公因子的含义，我们进一步地作正交旋转，其结果列于表 3.3。经过正交旋转后我们可看出：

(1) 公因子 F_1 中，男、女年轻人(20 岁以下)的载荷较高，因此 F_1 是反映"男女年轻人嗜好"因子．其贡献率为 38.9%．

(2) 公因子 F_2 中，女性成年人(20 岁以上)的载荷较高(绝对值)，F_2 解释为"女成年人嗜好"因子．其贡献率为 23.5%．

(3) 公因子 F_3 中，男性成年人(20 岁以上)的载荷较高，F_3 解释为"男成年人嗜好"因子，其贡献率为 28.9%．表 3.3 中的 $x_i(i=1,2,\cdots,10)$ 的含义如例 2.2 所示．

下面再列举一个 Q 型因子分析的例子，主要说明 Q 型因子分析在分类中的应用．

例 3.3 震旦一寒武纪地层内某矿床处于构造带中，由于构造复杂，使地层划分、时代归属等问题的看法不一．矿床中矿化岩石种类繁多，有泥板岩类、白云岩类、含砾泥板岩、辉绿岩等等，样品

表 3.4　27个样品在两个公因子上的载荷值

样品 \ 载荷 \ 因子	F_1	F_2	岩石类别
1	0.950	0.307	I
2	0.930	0.366	I
3	0.921	0.377	I
4	0.870	0.483	I
5	0.862	0.504	I
6	0.871	0.487	I
7	0.963	0.376	I
8	0.938	0.316	I
9	0.936	0.345	I
10	0.923	0.363	I
11	0.916	0.395	I
12	0.840	0.540	I
13	0.955	0.292	I
14	0.950	0.311	I
15	0.940	0.331	I
16	0.705	0.706	II_1
17	0.642	0.764	II_1
18	0.492	0.866	II_1
19	0.185	0.980	II_2
20	0.720	0.679	II_1
21	0.678	0.732	II_1
22	0.640	0.761	II_1
23	0.613	0.766	II_1
24	0.571	0.817	II_1
25	0.453	0.886	II_1
26	0.257	0.962	II_2
27	0.160	0.980	II_2

中各种岩石类型不易区分，如何进一步划分这些样品的类型是必须解决的实际问题，因为它是划分地层搞清构造的基础．为此采集了 27 个样品，每个样品测试了 SiO_2，Fe_2O_3，TiO_2 等 13 项指标，对 27 个样品进行 Q 型因子分析，希望通过对样品起共同影响

表 3.5 第 7 号和 27 号样品的指标数据

指标 \ 样号	7	27
SiO_2	30.5	5.30
Fe_2O_3	0.37	1.37
FeO	2.41	1.18
Al_2O_3	5.35	1.69
TiO_2	0.24	0.11
MnO	0.11	0.64
CaO	22.1	31.1
MgO	7.0	12.8
P_2O_5	0.04	0.10
$S_{全}$	0.12	0.83
V_2O_5	0.01	/
$C_{有}$	0.57	0.02
X	0.01	0.55

的公因子来对27个样品进行分类.

根据 27 个样品的 13 项指标测试数据(原始数据从略)计算两两样品之间的夹角余弦，得相似系数矩阵 Q，并计算其特征根为 $\lambda_1 = 23.927$, $\lambda_2 = 2.869$, $\lambda_3 = 0.069$, $\lambda_4 = 0.062, \cdots, \lambda_{27} = 0$, 这 27 个特征根的前两个已占总和的 99.4%，因此选两个公因子. 经正交旋转后 27 个样品在这两个公因子上的载荷列在表 3.4.

从表 3.4 可看出在公因子 F_1 上载荷最大的是第 7 号样品，其载荷值为 0.963，把第 7 号样品视为 F_1 所代表的地质作用下的典型样品，同样第 27 号样品是 F_2 所代表的地质作用下的典型样品. 我们将第 7 号样品与第 27 号样品的化学组成列在表 3.5，显然它们的化学组成有很大区别，尤其在 SiO_2, Al_2O_3, CaO, MgO 等的含量上，这两个样品代表了不同的成岩类型，其它样品可看作这两个样品的过渡. 将两个公因子作为坐标轴，把每个样品的两个载荷值作为坐标进行点图，如图 3.1 所示，从图中不仅可清晰看

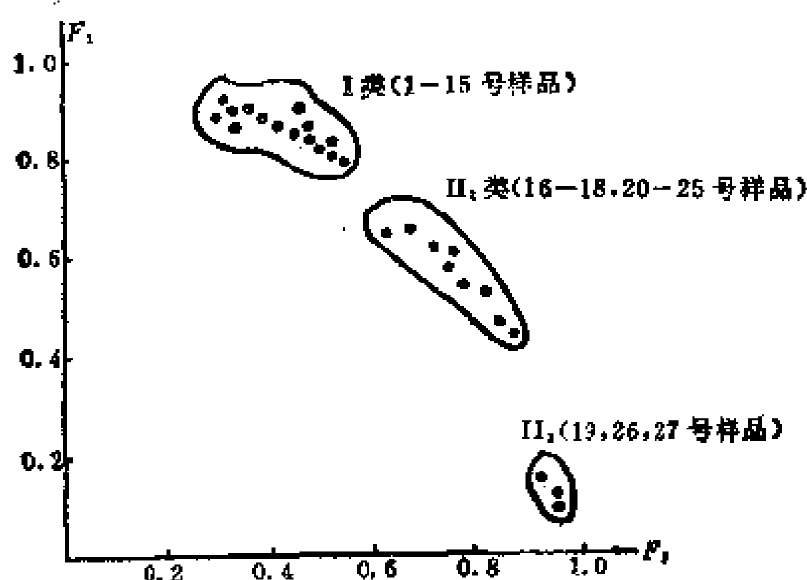


图 3.1　27 个样品按两个公因子的载荷分类

出如何分类，同时也可看出各样品间的过渡，从序号 1～15 为 I 类(为硅质泥质白云岩)，代表一种成岩作用；第 16～27 号为 II 类，结合化学成分，在 II 类中还可进一步划分为 II_1 类(为含硅质白云岩)和 II_2 类(白云岩)。II 类代表了另一种成岩作用。

(二)　因子得分

因子模型是 $\boldsymbol{x}-\boldsymbol{\mu}=\boldsymbol{A}\boldsymbol{F}+\boldsymbol{\epsilon}$. 虽然我们感兴趣的是模型中的载荷矩阵$\boldsymbol{A}$的估计及对公因子的解释，但有时要把公因子表示成变量的线性组合，对每个样品计算它们的公因子值，即因子得分。在许多实际问题中，例如地质、生物、气象等领域中常常要将样品根据地质作用、环境因素(公因子往往被解释为这些作用或因素）进行分类而这种分类可由因子得分来实现。下面我们将介绍用加权最小二乘法估计因子得分。

对于因子模型 $\boldsymbol{x}-\boldsymbol{\mu}=\boldsymbol{A}\boldsymbol{F}+\boldsymbol{\epsilon}$，先假设均值向量 $\boldsymbol{\mu}$，载荷矩阵$\boldsymbol{A}$和特殊因子方差都是已知的，而把特殊因子 $\boldsymbol{\epsilon}$ 看作误差，因为 $\mathrm{var}\varepsilon_i=\phi_i^2$，$(i=1,2,\cdots,p)$ 且不一定全等，因此采用加权最小二乘法来估计。在因子模型两边左乘 $\phi^{-\frac{1}{2}}$ 得

$$\phi^{-\frac{1}{2}}(x-\mu)=(\phi^{-\frac{1}{2}}A)F+\phi^{-\frac{1}{2}}\varepsilon$$

并记为

$$x^{*}=A^{*}F+\varepsilon^{*} \tag{3.9}$$

其中 $x^{*}=\phi^{-\frac{1}{2}}(x-\mu)$, $A^{*}=\phi^{-\frac{1}{2}}A$, $\varepsilon^{*}=\phi^{-\frac{1}{2}}\varepsilon$, 此时有

$$E\varepsilon^{*}=E\phi^{-\frac{1}{2}}\varepsilon=\phi^{-\frac{1}{2}}E\varepsilon=0$$

$$\mathrm{cov}(\varepsilon^{*})=E(\varepsilon^{*}\varepsilon^{*\prime})=E[\phi^{-\frac{1}{2}}\varepsilon\varepsilon'\phi^{-\frac{1}{2}}]=I$$

这就说明了(3.9)式为经典的回归模型，因此由经典回归的最小二乘法知 F 的估计为

$$\begin{aligned}\hat{F}&=(A^{*\prime}A^{*})^{-1}A^{*\prime}x^{*}=(A'\phi^{-\frac{1}{2}}\phi^{-\frac{1}{2}}A)^{-1}A'\phi^{-\frac{1}{2}}\phi^{-\frac{1}{2}}(x-\mu)\\&=(A'\phi^{-1}A)^{-1}A'\phi^{-1}(x-\mu)\end{aligned} \tag{3.10}$$

实际上,(3.10)式中 A,ϕ 和 μ 都是未知的，一种自然的想法是用它们的某种估计 $\hat{A},\hat{\phi}$ 和 $\hat{\mu}$ 来代替,在这里我们用

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i,$$

$\hat{A}$ 为正交旋转后的载荷矩阵，$\hat{\phi}=\mathrm{diag}(1-\hat{h}_1^2,\cdots,1-\hat{h}_p^2)$ 分别代替 μ,A 和 ϕ。于是对于样品 $x_j(j=1,2,\cdots,n)$ 的因子得分为

$$\hat{F}_j=(\hat{A}'\hat{\phi}^{-1}\hat{A})^{-1}\hat{A}'\hat{\phi}^{-1}(x_j-\hat{\mu}) \tag{3.11}$$

这里的 $\hat{F}_j$ 是 $m\times 1$ 维向量.

*§ 10.4　斜交因子模型及其解

前面讨论的因子模型中公因子是正交的，进行因子旋转也是正交旋转，但实际中对变量(或样品)都发生影响的公因子之间往往是相互联系的,即公因子之间是相关的,称这种相关的公因子为**斜交公因子**(简称斜交因子)，大量实验数据证明了斜交因子是普遍的,而正交因子只是在少数范围内存在,或者作为斜交因子的一种近似，本节是在正交因子的基础上来介绍斜交因子概念及求解方法.

（一） 斜交因子

当忽略了特殊因子后，且设 x 已标准化，正交因子模型可表为 $x = AF$，其中载荷矩阵 A 的元素 a_{ij} 刚好为 x_i 与 F_j 的相关系数。如果公因子之间不是正交的，即公因子之间是相关的．斜交公因子将记为 T，相应的载荷用 W 表示。当忽略了特殊因子后，斜交因子模型可写为

$$x = WT \tag{4.1}$$

或者写为

$$\left.\begin{aligned} x_1 &= w_{11}T_1 + w_{12}T_2 + \cdots + w_{1m}T_m \\ x_2 &= w_{21}T_1 + w_{22}T_2 + \cdots + w_{2m}T_m \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_p &= w_{p1}T_1 + w_{p2}T_2 + \cdots + w_{pm}T_m \end{aligned}\right\} \tag{4.2}$$

上式中 $T_1, \cdots, T_m$ 为斜交公因子，它们可视为斜坐标的单位向量，x_i 视为斜坐标系上的向量 p_i，而 w_{ij} 则为 p_i 在斜因子轴 T_j 上的坐标，称它为斜交因子载荷，与正交因子的情形不同，这里 w_{ij} 并不是 x_i 与 T_j 的相关系数。事实上，x_i 与 T_j 的相关系数

$$\begin{aligned} r_{x_i,T_j} = E(x_i \cdot T_j) &= E[(w_{i1}T_1 + \cdots + w_{ij}T_j + \cdots \\ &\quad + w_{im}T_m) \cdot T_j] = w_{i1}E(T_1T_j) + \cdots \\ &\quad + w_{ij}E(T_j,T_j) + \cdots + w_{im}E(T_m,T_j) \\ &\neq w_{ij} \end{aligned}$$

上式中的不等号成立是因为公因子相关的，所以 $E(T_i,T_j) \neq 0$。

将 $S = (r_{x_iT_j})_{p\times m}$ 称为**结构矩阵**，那么在正交因子模型之下，结构矩阵 S 与载荷矩阵 A 是一回事，而在斜交的情况下，结构矩阵 S 与载荷矩阵 W 就不是一回事。

由于斜交因子是相关的，记它们的相关系数矩阵为 U，其元素 u_{ij} 表示斜因子轴 T_i 与 T_j 的相关系数，而 T_i 与 T_j 为单位向量，由本章第二节(五)中知道在某个直角坐标系中 T_i 与 T_j 的夹角余弦就是它们的相关系数。我们取这个直角坐标系是由正交

因子轴 $F_1,\cdots,F_m$ 构成的,那么 T_i 在 $F_1,\cdots,F_m$ 上的坐标刚好就是 T_j 与它们的夹角余弦,即

$$t_{ij}=\|T_j\|\cos\theta_{F_i,T_j}=\cos\theta_{F_i,T_j}(i,j=1,2,\cdots,m) \quad (4.3)$$

$$T_j=[t_{1j},t_{2j},\cdots,t_{mj}]'(j=1,2,\cdots,m) \quad (4.4)$$

且

$$\sum_{i=1}^{m}t_{ij}^2=1 \quad (j=1,2,\cdots,m) \quad (4.5)$$

将(4.4)写成矩阵形式

$$T=[T_1,T_2,\cdots,T_m]=(t_{ij})_{m\times m} \quad (4.6)$$

这个矩阵 $T=(t_{ij})_{m\times m}$ 便是正交因子 $F_1,\cdots,F_m$ 变换成斜交因子的变换矩阵. 而斜交因子 T_i 与 T_j (单位向量)的相关系数就是 T_i 与 T_j 的夹角余弦:

$$\cos(\widehat{T_iT_j})=r_{T_iT_j}=\frac{T_i'T_j}{\|T_i\|\|T_j\|}$$

$$=T_i'T_j=\sum_{s=1}^{m}t_{si}t_{sj}=u_{ij} \quad (4.7)$$

$$i,j=1,2,\cdots,m.$$

从而斜交因子的相关系数矩阵为

$$U=T'T \quad (4.8)$$

设 $F_1,\cdots,F_m$ 为标准化变量 $x_i(i=1,2,\cdots,p)$ 的正交公因子,将 x_i 视为以 $F_1,\cdots,F_m$ 为坐标系中的一个单位向量, 记为 P_i, 它在该坐标系中的坐标就是 x_i 在各个公因子上的载荷, 即 $P_i=(a_{i1},\cdots,a_{im})'$. 而斜交因子 $T_j(j=1,2,\cdots,m)$ 也是该坐标系中的 m 个单位向量,由(4.4)知道其坐标为 $t_{1j},\cdots,t_{mj}$.在正交因子的坐标系下, x_i 与 T_j 的相关系数就是 P_i 与 T_j 的夹角余弦,即

$$r_{x_i,T_j}=\text{cov}(\widehat{P_iT_j})=\frac{P_i'T_j}{\|P_i\|\|T_j\|}=P_i'T_j=\sum_{s=1}^{m}a_{is}t_{sj} \quad (4.9)$$

写成矩阵形式为

$$S=(r_{x_iT_j})_{p\times m}=AT \quad (4.10)$$

其中A为正交因子载荷阵，T为(4.6)所定义。

有了上述的准备工作，下面要讨论如何求解斜交因子模型中的斜交载荷矩阵 W。斜交因子模型(4.2)为(略去了特殊因子)

$$x_i=w_{i1}T_1+\cdots+w_{im}T_m \quad (i=1,2,\cdots,p)$$

用 T_j 右乘上式两边，并求期望得

$$E(x_iT_j)=w_{i1}E(T_1T_j)+\cdots+w_{ij}E(T_jT_j)+\cdots+w_{im}E(T_m,T_j)$$

$$i=1,2,\cdots,p;\ j=1,2,\cdots,m.$$

注意到 x_i 与 T_j 都假设标准化，故上式可写为

$$r_{x_jT_j}=w_{i1}u_{1j}+w_{i2}u_{2j}+\cdots+w_{ij}+\cdots+w_{im}u_{mj} \quad (4.11)$$

写成矩阵形式为

$$S=WU \quad (4.12)$$

将(4.8)和(4.10)代入(4.12)得

$$AT=WT'T$$

从而有

$$W=AT(T'T)^{-1} \quad (4.13)$$

(4.13)式告诉我们只要知道正交因子载荷矩阵A及(4.6)式所定义的T矩阵就可以求出斜交载荷矩阵 W。

至此，对于斜交因子解，它涉及到三方面的内容：

(1) 斜交因子的相关系数矩阵：$U=(u_{ij})_{m\times m}=T'T$，

(2) 因子结构矩阵（变量与斜交因子的相关系数矩阵）：$S=AT$；

(3) 斜交因子载荷矩阵：$W=AT(T'T)^{-1}$。

无论要求 U、S、W中的那一个都必须求出 T，而由T的定义知道，它的第 j 列是斜交因子轴 T_j 在正交因子轴 $F_1,\cdots,F_m$ 的坐标系中的方位余弦。 如何求出T呢？ 首先我们引进斜交参考解。

(二) 斜交参考解

仍在正交因子轴 $F_1,\cdots,F_m$ 构成的坐标系中讨论，斜交因子轴 $T_1,\cdots,T_m$，是该坐标系中的 m 个单位向量。在该坐标系中再引入 m 个单位向量：$\Lambda_1,\Lambda_2,\cdots,\Lambda_m$。称它们为斜交参考轴。且要求 Λ_i 垂直于由 $T_1,\cdots,T_{i-1},T_{i+1},\cdots,T_m$ 构成的超平面。当 $m=3$ 时，Λ_1 是 T_2,T_3 构成的平面法线，Λ_2 及 Λ_3 分别为 T_1,T_3 和 T_1,T_2 构成的平面法线。当 $m=2$ 时，两个斜交参考轴 Λ_1,Λ_2 分别垂直于 T_2 和 T_1，如图 4.1 所示。若点群(1～6)在 T_1 附近，那么在 Λ_2 的投影接近于零；若点群(7～11)在 T_2 附近，那么在 Λ_1 的投影接近于零。

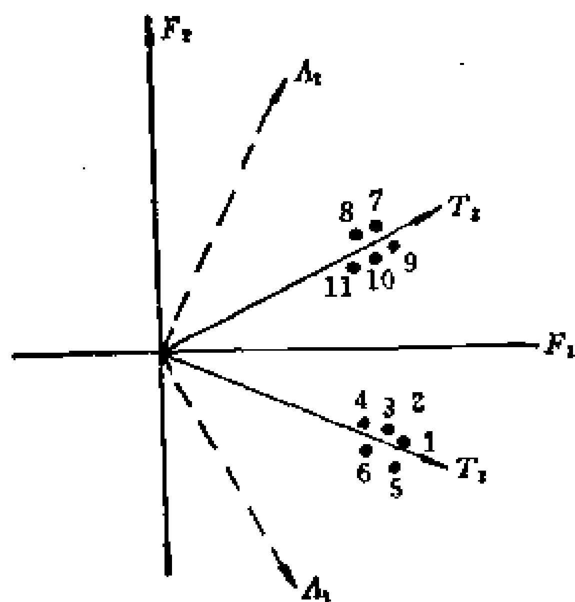


图 4.1

我们将 $T_1,\cdots,T_m$ 和 $\Lambda_1,\cdots,\Lambda_m$ 视为两组斜交因子，我们已知道(4.6)式的 $T=(t_{ij})_{m\times m}$ 是正交因子轴变换到斜交因子 $T_1,\cdots,T_m$ 的变换矩阵，它的元素 t_{ij} 刚好为 T_i 与 F_j 的夹角余弦，类似地设 $\Lambda=(\lambda_{ij})_{m\times m}$ 是正交因子 $F_1,\cdots,F_m$ 变换到斜交参考轴 $\Lambda_1,\cdots,\Lambda_m$ 的变换矩阵，其中 λ_{ij} 是 Λ_i 与 F_j 的夹角余弦，现在我们来找出 T 与 Λ 之间的关系。令

$$D=T'\Lambda=\begin{pmatrix}T_1'\\ \vdots\\ T_m'\end{pmatrix}(\Lambda_1,\cdots,\Lambda_m)$$

$$=\begin{pmatrix}T_1'\Lambda_1 & T_1'\Lambda_2 & \cdots & T_1'\Lambda_m\\ T_2'\Lambda_1 & T_2'\Lambda_2 & \cdots & T_2'\Lambda_m\\ \cdots & \cdots & \cdots & \cdots\\ T_m'\Lambda_1 & T_m'\Lambda_2 & \cdots & T_m'\Lambda_m\end{pmatrix} \tag{4.14}$$

由于 T_i 与 $\Lambda_j(i \neq j)$ 垂直，因此 $T_i'\Lambda_j = 0\ (i \neq j)$，因而有

$$D = \mathrm{diag}(T_1'\Lambda_1, \cdots, T_m'\Lambda_m) \tag{4.15}$$

设 $\Lambda^{-1} = (\mu_{ij})_{m\times m}$，则有

$$T' = D\Lambda^{-1} = \begin{pmatrix} T_1'\Lambda_1\mu_{11} & T_1'\Lambda_1\mu_{12} & \cdots & T_1'\Lambda_1\mu_{1m} \\ T_2'\Lambda_2\mu_{21} & T_2'\Lambda_2\mu_{22} & \cdots & T_2'\Lambda_2\mu_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ T_m'\Lambda_m\mu_{m1} & T_m'\Lambda_m\mu_{m2} & \cdots & T_m'\Lambda_m\mu_{mm} \end{pmatrix} \tag{4.16}$$

又因为 T 中的每列元素的平方和为 1，即

$$(T_i'\Lambda_i)^2(\mu_{i1}^2 + \cdots + \mu_{im}^2) = 1$$

所以有

$$T_i'\Lambda_i = \left(\sum_{j=1}^{m} \mu_{ij}^2\right)^{-\frac{1}{2}}\ (i = 1, 2, \cdots, m.) \tag{4.17}$$

将(4.17)代人(4.16)得

$$T' = \begin{pmatrix} \mu_{11}\Big/\left(\sum_{j=1}^{m} \mu_{1j}^2\right)^{\frac{1}{2}} & \mu_{12}\Big/\left(\sum_{j=1}^{m} \mu_{1j}^2\right)^{\frac{1}{2}} & \cdots & \mu_{1m}\Big/\left(\sum_{j=1}^{m} \mu_{1j}^2\right)^{\frac{1}{2}} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{m1}\Big/\left(\sum_{j=1}^{m} \mu_{mj}^2\right)^{\frac{1}{2}} & \mu_{m2}\Big/\left(\sum_{j=1}^{m} \mu_{mj}^2\right)^{\frac{1}{2}} & \cdots & \mu_{mm}\Big/\left(\sum_{j=1}^{m} \mu_{mj}^2\right)^{\frac{1}{2}} \end{pmatrix} \tag{4.18}$$

由(4.18)可以看出 T' 就是 Λ^{-1} 按行"单位化"而得到的矩阵，求 T 的问题就归结为求 Λ。根据我们引进斜交参考轴 $\Lambda_1, \cdots, \Lambda_m$ 的特点，目前已有一种近似求 Λ 的 Promax 斜交旋转的方法。其步骤为：

(i) 按 § 10.2 的方法求出初始载荷矩阵 A；

(ii) 将 A 按行单位化得 A^*；

(iii) 将 A^* 的各元素取绝对值的 K 次幂 ($K \geqslant 2$) 并保留原来的符号得矩阵 H；

(iv) 由已求出的 A^*, H 两矩阵，考虑模型

$$H = A^*C \tag{4.19}$$

表 4.1　十个变量的相关系数矩阵

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	1									
x_2	0.5842	1								
x_3	0.5629	0.9825	1							
x_4	0.2906	0.5806	0.6789	1						
x_5	0.4735	0.8813	0.9251	0.8197	1					
x_6	0.2153	0.4243	0.4943	0.7954	0.7233	1				
x_7	0.2002	0.3959	0.4618	0.7604	0.6946	0.9927	1			
x_8	0.0633	0.0198	0.0279	0.6214	0.2554	0.6581	0.6282	1		
x_9	0.5001	0.7139	0.7846	0.6432	0.7479	0.4073	0.3840	0.0007	1	
x_{10}	0.2892	0.3445	0.4053	0.4690	0.3959	0.4216	0.3823	0.3678	0.3079	1

表 4.2 初始载荷矩阵 A

载荷 公因子 / 变量	F_1	F_2	F_3	F_4	F_5	共同度
x_1	0.5305	−0.4774	0.3375	0.6075	0.0801	0.9988
x_2	0.8182	−0.4899	−0.0636	−0.0813	−0.1748	0.9507
x_3	0.8785	−0.4203	−0.0560	−0.1308	−0.0874	0.9763
x_4	0.8933	0.2611	−0.0733	−0.1011	0.2473	0.9430
x_5	0.9535	−0.1407	−0.1685	−0.0643	−0.0851	0.9687
x_6	0.8161	0.4943	−0.1348	0.1324	−0.1661	0.9737
x_7	0.7866	0.4983	−0.1665	0.1593	−0.1950	0.9582
x_8	0.4158	0.8136	0.1161	0.0757	0.2279	0.9060
x_9	0.7613	−0.3890	−0.0970	−0.1303	0.3839	0.9046
x_{10}	0.5494	0.1576	0.7511	−0.2819	−0.1103	0.9886
公因子贡献	57.7%	20.6%	14%	4.5%	3.9%	
累积贡献	57.7%	78.3%	87.3%	91.8%	95.7%	

表 4.3 因子结构矩阵 S

变量	结构				
	T_1	T_2	T_3	T_4	T_5
x_1	0.0379	0.0343	0.0021	0.8080	0.0288
x_2	0.7511	−0.0573	0.0629	0.0303	0.0007
x_3	0.7531	−0.0379	0.0797	−0.0056	0.0985
x_4	0.3188	0.3820	0.0157	−0.0570	0.3777
x_5	0.6329	0.2469	−0.0208	−0.0319	0.0918
x_6	0.1028	0.7643	−0.0293	0.0123	−0.0865
x_7	0.0867	0.7866	−0.0635	0.0176	−0.1213
x_8	−0.3765	0.6630	0.0981	0.0237	0.2044
x_9	0.6234	−0.1676	−0.0813	0.0448	0.5373
x_{10}	0.0861	−0.0123	0.8534	0.0020	−0.0503

求 C 的最小二乘估计，易知 C 的最小二乘估计为

$$C = (A^{*\prime}A^{*})^{-1}A^{*\prime}H$$

(v) 将 C 按列单位化，得斜交参考矩阵 Λ；

(vi) 将 Λ^{-1} 按行单位化，得斜交变换矩阵 T'；

(vii) 按(4.8)，(4.10)，(4.13) 便可求出斜交因子的相关系数矩阵 U，结构矩阵 S 及斜交载荷阵 W。

在进行 Promax 旋转时，可让 K 取 2，3，4，⋯ 依次进行，比较各次的结果，当斜交因子相关矩阵 U 相对稳定下来就停止，由经验一般 $K = 2 \sim 4$。如果给定的一组变量之间的相关关系越复杂 K 值就越大。

关于斜交因子解的详细讨论可参看[34]第七章第五节。

例 4.1 应用因子分析研究某地区含铂超基性岩中铂族元素的赋存状态和聚集规律，为本区进一步的研究工作提供依据。对该区的 90 个样品分析了 10 个指标：x_1 = Au，x_2 = Pd，x_3 = Pt，x_4 = Ir，x_5 = Rh，x_6 = Os，x_7 = Ru，x_8 = Cr_2O_3，x_9 = Ni，x_{10} = Co。由原始数据计算得相关系数矩阵列在表 4.1，并计算其

表 4.4 斜交载荷矩阵W

变量	斜交载荷				
	T_1	T_2	T_3	T_4	T_5
x_1	0.0486	0.0437	0.0024	0.9654	0.0327
x_2	0.9646	-0.0731	0.0728	0.0362	0.0008
x_3	0.9671	-0.0483	0.0924	-0.0067	0.1120
x_4	0.4094	0.4875	0.0182	-0.0681	0.4239
x_5	0.8127	0.3151	-0.0241	-0.0381	0.1044
x_6	0.1320	0.9753	-0.0340	0.0146	-0.0984
x_7	0.1113	1.0038	-0.0735	0.0210	-0.1379
x_8	-0.4835	0.8461	0.1136	0.0283	0.2323
x_9	0.9006	-0.2139	-0.0942	0.0536	0.6107
x_{10}	0.1105	-0.0157	0.9887	0.0024	-0.0571

特征根和相应单位化特征向量。当提取前 5 个公因子时，公因子的累积贡献率达 95.7%，前 5 个公因子已包含了原始变量的绝大部分信息，其初始载荷矩阵列在表 4.2。

对表 4.2 的初始载荷矩阵进行 Promax 斜旋转，这种斜旋转包括结构矩阵 S，斜载荷矩阵W和斜因子相关矩阵 U。分别计算的结果如下：

（1）因子结构矩阵 $S = AT$，列于表 4.3。因子结构矩阵中，等于零或接近于零的元素越多，说明斜结构越简单，由表 4.3 可看出 x_2 = Pd，x_3 = Pt，x_5 = Rh，x_9 = Ni 在斜因子轴 T_1 上的投影（即这几个变量与 T_1 的相关系数）较大，而在其它因子上的投影都较小；同样，x_6 = Os，x_7 = Ru，x_8 = Cr_2O_3 在斜因子轴 T_2 上的投影较大，而在其它因子轴上的投影较小，头两个斜交因子已基本包含了关于铂族元素的主要信息。

（2）斜交载荷矩阵 $W = AT(T'T)^{-1}$，列于表 4.4。从斜交载荷矩阵可以看出在第一个斜交公因子 T_1 上的载荷比较高的有 x_2 = Pd，x_3 = Pt，x_5 = Rh，x_9 = Ni，因此 T_1 是一个与镍矿化有关的铂族元素（Pd,Pt,Rh）的聚集过程。T_2 是一个与铬矿

化有关的铂族元素（Os,Ru）的聚集过程．此外，铂族元素Ir在 T_1 和 T_2 这两个聚集过程中有一定含量出现，但它的聚集过程还单独与镍的组合出现，这反映在公因子 T_5 上。至于公因子 T_3 与 T_4 则分别反映了 Co 与 Au．它们与铂族元素的关系不大。

斜交因子载荷矩阵是 Promax 斜旋转的主要内容，根据斜交因子模型，可将各个变量近似地表示成斜交因子的线性组合，以变量 x_2 = Pd 为例，有

$$x_2 = 0.9646T_1 - 0.0731T_2 + 0.0728T_3 + 0.0362T_4 + 0.0008T_5$$

可见，Pd 几乎由 T_1 所提供，即 Pd 是由镍矿化的铂族元素聚集过程所提供。

(3) 斜交因子相关矩阵 $U = T'T$，列于表 4.5。T_1 与 T_2 的

表 4.5 斜交因子相关矩阵 U

	T_1	T_2	T_3	T_4	T_5
T_1	1				
T_2	0.3956	1			
T_3	0.2310	0.4338	1		
T_4	0.5087	0.0766	0.1997	1	
T_5	0.0971	0.4206	0.3634	0.0447	1

相关系数为 0.3956，T_2 与 T_5 的相关系数为 0.4206 皆有一定的显著性，说明了 T_1 所代表的 Pd,Pt,Rh 的镍矿化与 T_2 所代表的 Os,Ru 的铬矿化存在一定的正相关关系，同样铬矿化与部分镍铱矿化也存在一定的正相关关系。

最后，在结束本节时我们要指出，因子分析虽然具有很大的直观吸引力，但在对公因子进行解释时是带有很强的专业性、经验性，甚至于还带有某种主观性。不少的应用实例还表明了因子分析并没有给出满意的结果。因子分析也没有一个定量的方法去判断因子分析结果的优劣性。一般说来，如果研究者仔细地考察了

所得的结果，满意地解释了公因子的意义并获得较一致的看法，那么这项应用就可以认为是成功的.

*§ 10.5　对应分析方法

(一)　问题的提出

因子分析分为R型和Q型，寻找变量的公因子就采用R型，寻找样品的公因子就采用Q型. R型是从变量的相关系数矩阵（或协方差阵)出发，Q型是从样品的相似矩阵出发. 在因子分析中把R型和Q型互相割裂单独进行，有些问题只作R型分析，有些只作Q型分析，即使有些问题同时作了两种分析，在解释时也无法将它们有机地联系起来. 然而变量和样品是分不开的，我们要通过样品来获得变量的观测值，反之又要通过变量值来对样品进行刻画和解释，即变量之间的关系只有通过样品来体现，样品的特征和相似性又必须通过变量来刻画，这说明R型分析与Q型分析是不可分割的.

此外，一般在实际的问题中，所取的样品数目远远地大于变量的数目，这样给Q型因子分析带来计算上的困难，譬如有 150 个样品，每个样品分析 10 个变量，如果采用主成分法对载荷矩阵进行估计作R型因子分析时只须计算 10×10 阶的变量相关系数矩阵的特征根和特征向量，而Q型因子分析则要计算 150×150 阶的样品相似矩阵的特征根和特征向量. 这个计算量是十分可观的.

除了上述原因，在因子分析中还存在这样的问题：为了能够将量纲不同的变量进行比较往往对变量进行标准化处理，然而这种标准化只能对变量进行，对样品则无从谈标准化，所以标准化对变量和样品是非对等的，这也给R型和Q型因子分析之间的联系带来障碍.

针对以上问题，1970 年 Beozecri 提出了对应分析方法，它综合了R型和Q型因子分析的优点，并将它们统一起来使得由R型的分析结果很容易得到Q型的分析结果，这就克服了Q型分析计

算量大的困难；更重要的是可以把变量和样品的载荷反映在相同的公因子轴上，这样就把变量和样品联系起来便于解释和推断。例如在气象问题中研究各种气象因素和样品类型的关系时，如果我们用头两个因子轴作一个坐标系，将变量和样品的载荷同时在坐标系上点图，相邻近的样品点属于同一类型，而这一类型附近的变量点可以解释这一类样品的成因，这对于了解气象过程是很有益的。在地质问题中也常常需要用这种方法来解释地质作用或过程。

从以上的讨论，对应分析的关键所在是使 R 型和 Q 型分析中变量点和样品点的载荷能反映在相同的公因子轴上，这就要求从 R 型的相关系数矩阵出发与从 Q 型分析的相似矩阵出发，导出的非零特征根应该相同，这些特征根正是各自公因子提供的方差贡献，因此我们可以用相同的因子轴去同时表示变量与样品。

（二） 对应分析的数据变换方法

设有 n 个样品，每个样品有 p 个变量，即资料阵为

$$X=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix}$$

对 X 的元素 x_{ij} 要求都大于 0（否则，对所有的数据同加一个数使其满足大于 0）。对资料阵按行、列求和，并求出总和，即

		样品				行和
		x_1	x_2	$\cdots$	x_n	
变量	x_1	x_{11}	x_{12}	$\cdots$	x_{1n}	$x_{1\cdot}$
	x_2	x_{21}	x_{22}	$\cdots$	x_{2n}	$x_{2\cdot}$
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	x_p	x_{p1}	x_{p2}	$\cdots$	x_{pn}	$x_{p\cdot}$
列和		$x_{\cdot 1}$	$x_{\cdot 2}$	$\cdots$	$x_{\cdot n}$	T

(5.1)

其中

$$x_{\cdot j}=\sum_{i=1}^{p}x_{ij},\quad x_{i\cdot}=\sum_{j=1}^{n}x_{ij},$$

$$T=\sum_{i=1}^{p}\sum_{j=1}^{n}x_{ij}=\sum_{i=1}^{p}x_{i\cdot}=\sum_{j=1}^{n}x_{\cdot j}.$$

用总和T去除每一个 x_{ij} 得到

$$V=\frac{1}{T}X=\begin{pmatrix}v_{11} & v_{12} & \cdots & v_{1n}\\ v_{21} & v_{22} & \cdots & v_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ v_{p1} & v_{p2} & \cdots & v_{pn}\end{pmatrix}\tag{5.2}$$

不难看出 $0<v_{ij}<1$，且 $\sum_i\sum_j v_{ij}=1$，因而 v_{ij}可解释为“概率”。

从矩阵V出发，先考虑R型分析。对V矩阵的每一列求和，即 $v_{\cdot j}=\sum_{i=1}^{p}v_{ij}$，$(j=1,2,\cdots,n)$，然后用 $v_{\cdot j}$ 去除第j列的每个元素，得到矩阵

$$\begin{pmatrix}\frac{v_{11}}{v_{\cdot 1}} & \frac{v_{12}}{v_{\cdot 2}} & \cdots & \frac{v_{1n}}{v_{\cdot n}}\\ \frac{v_{21}}{v_{\cdot 1}} & \frac{v_{22}}{v_{\cdot 2}} & \cdots & \frac{v_{2n}}{v_{\cdot n}}\\ \vdots & \vdots & \vdots & \vdots\\ \frac{v_{p1}}{v_{\cdot 1}} & \frac{v_{p2}}{v_{\cdot 2}} & \vdots & \frac{v_{pn}}{v_{\cdot n}}\end{pmatrix}\tag{5.3}$$

将每一列视为p维空间中的一个样品点，每个样品点的坐标是各个变量在该样品中的相对比例，经过这个变换后对n个样品点的研究变成了对n个样品点的相对关系的研究，对(5.3)中的样品点之间的距离采用加权距离，第α个样品与第β个样品加权距离公式为

$$D(\alpha,\beta)=\left[\sum_{i=1}^{p}\frac{1}{v_{i\cdot}}\left(\frac{v_{i\alpha}}{v_{\cdot\alpha}}-\frac{v_{i\beta}}{v_{\cdot\beta}}\right)^2\right]^{\frac{1}{2}}$$

$$=\left[\sum_{i=1}^{p}\left(\frac{v_{i\alpha}}{\sqrt{v_{i\cdot}v_{\cdot\alpha}}}-\frac{v_{i\beta}}{\sqrt{v_{i\cdot}v_{\cdot\beta}}}\right)^2\right]^{\frac{1}{2}} \tag{5.4}$$

为了沿用普通的欧氏距离，今后假定第 j 个样品点具有如下的坐标:

$$\left(\frac{v_{1j}}{\sqrt{v_{1\cdot}v_{\cdot j}}},\ \frac{v_{2j}}{\sqrt{v_{2\cdot}v_{\cdot j}}},\ \cdots,\frac{v_{pj}}{\sqrt{v_{p\cdot}\ v_{\cdot j}}}\right)'\ (j=1,2,\cdots,n) \tag{5.5}$$

从而我们可以得到一个新的矩阵

$$V^*=\begin{pmatrix}\dfrac{v_{11}}{\sqrt{v_{1\cdot}v_{\cdot1}}}, & \dfrac{v_{12}}{\sqrt{v_{1\cdot}v_{\cdot2}}}, & \cdots, & \dfrac{v_{1n}}{\sqrt{v_{1\cdot}v_{\cdot n}}}\\ \dfrac{v_{21}}{\sqrt{v_{2\cdot}v_{\cdot1}}} & \dfrac{v_{22}}{\sqrt{v_{2\cdot}v_{\cdot2}}} & \cdots & \dfrac{v_{2n}}{\sqrt{v_{2\cdot}v_{\cdot n}}}\\ \cdots & \cdots & \cdots & \cdots\\ \dfrac{v_{p1}}{\sqrt{v_{p\cdot}\ v_{\cdot1}}} & \dfrac{v_{p2}}{\sqrt{v_{p\cdot}\ v_{\cdot2}}} & \cdots & \dfrac{v_{pn}}{\sqrt{v_{p\cdot}\ v_{\cdot n}}}\end{pmatrix} \tag{5.6}$$

V^* 阵的每一列仍视为一个样本点，每一行仍视为一个变量的 n 个观测值。由 V^* 阵出发，计算各变量的协方差阵。给第 α 个样品以 $v_{\cdot\alpha}$ 的权求各个变量的加权平均，此时第 i 个变量的加权平均为

$$\sum_{\alpha=1}^{n}v_{\cdot\alpha}\cdot\frac{v_{i\alpha}}{\sqrt{v_{i\cdot}v_{\cdot\alpha}}}=\frac{1}{\sqrt{v_{i\cdot}}}\sum_{\alpha=1}^{n}v_{i\alpha}=\sqrt{v_{i\cdot}}$$

$$(i=1,2,\cdots,p).$$

同样地用加权的办法计算第 i 和第 j 个变量的协方差为

$$\sum_{\alpha=1}^{n}v_{\cdot\alpha}\left(\frac{v_{i\alpha}}{\sqrt{v_{i\cdot}\ v_{\cdot\alpha}}}-\sqrt{v_{i\cdot}}\right)\left(\frac{v_{j\alpha}}{\sqrt{v_{j\cdot}v_{\cdot\alpha}}}-\sqrt{v_{j\cdot}}\right)$$

$$=\sum_{\alpha=1}^{n}\left(\frac{v_{i\alpha}}{\sqrt{v_{i\cdot}}\sqrt{v_{\cdot\alpha}}}-\sqrt{v_{i\cdot}v_{\cdot\alpha}}\right)\left(\frac{v_{j\alpha}}{\sqrt{v_{j\cdot}}\sqrt{v_{\cdot\alpha}}}-\sqrt{v_{j\cdot}v_{\cdot\alpha}}\right)$$

$$= \sum_{\alpha=1}^{n} w_{i\alpha} w_{j\alpha} = a_{ij} \tag{5.7}$$

其中

$$w_{i\alpha} = \left(\frac{v_{i\alpha}}{\sqrt{v_{i\cdot}}\sqrt{v_{\cdot\alpha}}} - \sqrt{v_{i\cdot} v_{\cdot\alpha}} \right),$$

$$(i = 1, 2, \cdots, p, \quad \alpha = 1, 2, \cdots, n)$$

$$a_{ij} = \sum_{\alpha=1}^{n} w_{i\alpha} w_{j\alpha} \quad (i, j = 1, 2, \cdots, p)$$

写成矩阵形式有

$$W = \{w_{i\alpha}\}_{p\times n}$$

$$A = \{a_{ij}\}_{p\times p} = WW'$$

A称为资料阵(5.6)的变量样本协方差阵，但注意到

$$w_{i\alpha} = \frac{v_{i\alpha}}{\sqrt{v_{i\cdot}}\sqrt{v_{\cdot\alpha}}} - \sqrt{v_{i\cdot} v_{\cdot\alpha}} = \frac{v_{i\alpha} - v_{i\cdot} v_{\cdot\alpha}}{\sqrt{v_{i\cdot}}\sqrt{v_{\cdot\alpha}}}$$

$$= \frac{x_{i\alpha} - x_{i\cdot} x_{\cdot\alpha}/T}{\sqrt{x_{i\cdot} x_{\cdot\alpha}}} = z_{i\alpha} \tag{5.8}$$

从而

$$W = Z, \quad A = ZZ' \tag{5.9}$$

其中Z是$p \times n$阶矩阵，其元素由(5.8)所定义。

我们就从协方差阵A出发进行R型因子分析（稍后一点我们将会看到它与Q型分析的关系）。然后再来考虑Q型分析的情况。

类似于R型的做法，R型是对列进行变换处理，而在Q型中则对行作相对等的变换处理，这样便可得到与(5.6)相对应的资料阵

$$V^{**} = \begin{pmatrix} \frac{v_{11}}{\sqrt{v_{\cdot1} v_{1\cdot}}}, & \frac{v_{12}}{\sqrt{v_{\cdot2} v_{1\cdot}}}, & \cdots, & \frac{v_{1n}}{\sqrt{v_{\cdot n}\, v_{1\cdot}}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{v_{p1}}{\sqrt{v_{\cdot1}\, v_{p\cdot}}}, & \frac{v_{p2}}{\sqrt{v_{\cdot2}\, v_{p\cdot}}}, & \cdots, & \frac{v_{pn}}{\sqrt{v_{\cdot n}\, v_{p\cdot}}} \end{pmatrix} \tag{5.10}$$

仿照R型利用加权的方法可求出任意两个样品α及β间的协方差

$b_{\alpha\beta}$，写成矩阵形式：

$$B = \{b_{\alpha\beta}\}_{n\times n} \tag{5.11}$$

其中

$$\begin{aligned} b_{\alpha\beta} &= \sum_{i=1}^{p} v_{i\cdot}\left(\frac{v_{i\alpha}}{\sqrt{v_{\cdot\alpha}v_{i\cdot}}} - \sqrt{v_{\cdot\alpha}}\right)\left(\frac{v_{i\beta}}{\sqrt{v_{\cdot\beta}v_{i\cdot}}} - \sqrt{v_{\cdot\beta}}\right) \\ &= \sum_{i=1}^{p}\left(\frac{v_{i\alpha}}{\sqrt{v_{i\cdot}}\sqrt{v_{\cdot\alpha}}} - \sqrt{v_{i\cdot}}\sqrt{v_{\cdot\alpha}}\right) \\ &\quad \cdot\left(\frac{v_{i\beta}}{\sqrt{v_{i\cdot}}\sqrt{v_{\cdot\beta}}} - \sqrt{v_{i\cdot}}\sqrt{v_{\cdot\beta}}\right) \\ &= \sum_{i=1}^{p}\left(\frac{x_{i\alpha} - x_{i\cdot}x_{\cdot\alpha}/T}{\sqrt{x_{i\cdot}x_{\cdot\alpha}}}\right)\left(\frac{x_{i\beta} - x_{i\cdot}x_{\cdot\beta}/T}{\sqrt{x_{i\cdot}x_{\cdot\beta}}}\right) \\ &= \sum_{i=1}^{p} z_{i\alpha}z_{i\beta} \end{aligned}$$

从而得

$$B = Z'Z \tag{5.12}$$

由(5.9)与(5.12)可以看出 A 与 B 之间存在着简单的对应关系，由矩阵知识不难证明如下定理成立(习题 10.4)。

定理 5.1 设 λ_i 是 A 的非零特征根，e_i 为相应的特征向量，则有

(i) A 与 B 的所有非零特征根相等；

(ii) B 的非零特征根 λ_i 所对应的特征向量为 $Z'e_i$。

此定理告诉我们只须从 A 出发进行 R 型因子分析，就可容易地得到 Q 型因子分析的结果，另外 A 与 B 具有相同的非零特征根，注意到特征根是对应的公因子所提供的方差贡献这一事实，那么就可用相同的公因子轴去表示变量和样品。

（三） 对应分析的计算步骤及例

将(二)中讨论的具体实施步骤归纳如下：

设 $X=\{x_{ij}\}_{p\times n}$ 是 n 个样品的 p 个指标的观测资料阵，且 $x_{ij}>0$。

(i) 将 X 按行、列求和，并求出总和，并记行和为 $x_{i\cdot}=\sum_{j=1}^{n}x_{ij}$ $(i=1,2,\cdots,p)$，列和为 $x_{\cdot j}=\sum_{i=1}^{p}x_{ij}$，$(j=1,\cdots,n)$，总和 $T=\sum_{i=1}^{p}\sum_{j=1}^{n}x_{ij}=\sum_{i=1}^{p}x_{i\cdot}=\sum_{j=1}^{n}x_{\cdot j}$。

(ii) 计算矩阵 $Z=\{z_{ij}\}_{p\times n}$，其中

$$z_{ij}=\frac{x_{ij}-x_{i\cdot}x_{\cdot j}/T}{\sqrt{x_{i\cdot}x_{\cdot j}}}\ (i=1,2,\cdots,p,\ j=1,2,\cdots,n)$$

(iii) 进行 R 型因子分析。计算 $A=ZZ'$ 的特征根 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p$。按其累积百分比 $\left(\sum_{j=1}^{m}\lambda_j\Big/\sum_{j=1}^{p}\lambda_j\right)100\%\geqslant80\%$。取前 m 个特征根，即取 m 个公因子，计算其相应的单位化特征向量 $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_m$，从而得 R 型因子载荷阵：

	F_1	F_2	$\cdots$	F_m
x_1	$e_{11}\sqrt{\lambda_1}$	$e_{12}\sqrt{\lambda_2}$	$\cdots$	$e_{1m}\sqrt{\lambda_m}$
x_2	$e_{21}\sqrt{\lambda_1}$	$e_{22}\sqrt{\lambda_2}$	$\cdots$	$e_{2m}\sqrt{\lambda_m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_p	$e_{p1}\sqrt{\lambda_1}$	$e_{p2}\sqrt{\lambda_2}$	$\cdots$	$e_{pm}\sqrt{\lambda_m}$

并在两两因子轴平面上作变量的点图。

(iv) 进行 Q 型分析。对 (iii) 中 A 的前 m 个特征根计算相应于 B 矩阵的特征向量 $\boldsymbol{e}_i^*=Z'\boldsymbol{e}_i\ (i=1,2,\cdots,m)$，并将 $\boldsymbol{e}_i^*$ 单位化，从而得到样品的载荷矩阵：

	G_1	G_2	$\cdots$	G_m
x_1	$e_{11}^*\sqrt{\lambda_1}$	$e_{12}^*\sqrt{\lambda_2}$	$\cdots$	$e_{1m}^*\sqrt{\lambda_m}$
x_2	$e_{21}^*\sqrt{\lambda_1}$	$e_{22}^*\sqrt{\lambda_2}$	$\cdots$	$e_{2m}^*\sqrt{\lambda_m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	$e_{n1}^*\sqrt{\lambda_1}$	$e_{n2}^*\sqrt{\lambda_2}$	$\cdots$	$e_{nm}^*\sqrt{\lambda_m}$

在R型因子平面上作相应的样品点图。

(v) 根据具体问题进行解释和推断。

例 5.1 长江中下游六月份旱涝的对应分析。样本数 $n=29$。变量有 5 个，它们是：x_1 — 九月份亚洲地区经向环流指数，x_2 — 一月份亚欧地区月平均经向环流指数，x_3 — 十月份 85°N，150°E—170°W 五点 500 毫帕月平均高度的平均值减 500 位势什米；x_4 — 长江中下游十月份降水量，x_5 — 十二月份 50°N，180°E—160°W 三点 500 毫帕月平均高度的平均值减 500 位势什米。原始资料矩阵 $X=\{x_{ij}\}_{5\times 29}$（从略）。

(i) 由原始资料阵容易计算其行和 $x_{i\cdot}(i=1,2,\cdots,5)$，列和 $x_{\cdot j}$，$(j=1,2,\cdots,29)$ 以及总和 T。

(ii) 计算 Z 阵。Z 阵的元素按公式(5.8)计算(从略)。

(iii) 计算 $A=ZZ'$：

$$A=\begin{bmatrix} 0.00085 & & & & \\ 0.00076 & 0.00098 & & & \\ 0.00381 & 0.00040 & 0.03741 & & \\ -0.00547 & -0.00628 & -0.03691 & 0.05826 & \\ 0.00423 & 0.00484 & 0.02012 & -0.04844 & 0.04825 \end{bmatrix}$$

并计算A的特征根及相应的单位化特征向量，列于表 5.1。

前两个特征根已占特征根总和的 99.6%，故取前两个公因子。按 $e_1\sqrt{\lambda_1}$，$e_2\sqrt{\lambda_2}$ 计算得R型的因子分析载荷矩阵如表 5.2。

表 5.1　$A=ZZ'$ 的特征根及特征向量

特征根	$\lambda_1=0.1231$	$\lambda_2=0.0220$	$\lambda_3=0.0005$	$\lambda_4=0.0001$	$\lambda_5=0.0000$
	e_1	e_2	e_3	e_4	e_5
特征向量	−0.0646	−0.0213	0.7423	−0.6634	0.0656
	−0.0742	−0.0256	0.6543	0.7479	0.0795
	−0.4370	−0.7944	−0.1113	−0.0163	0.4065
	0.6879	−0.0060	−0.0071	−0.0029	0.7257
	−0.5711	0.6064	−0.0923	−0.0132	0.5454

表 5.2　R 型因子载荷阵

因子 变量	F_1	F_2
x_1	−0.0227	−0.0032
x_2	−0.0260	−0.0038
x_3	−0.1533	−0.1179
x_4	0.2414	−0.0009
x_5	−0.2004	0.0900

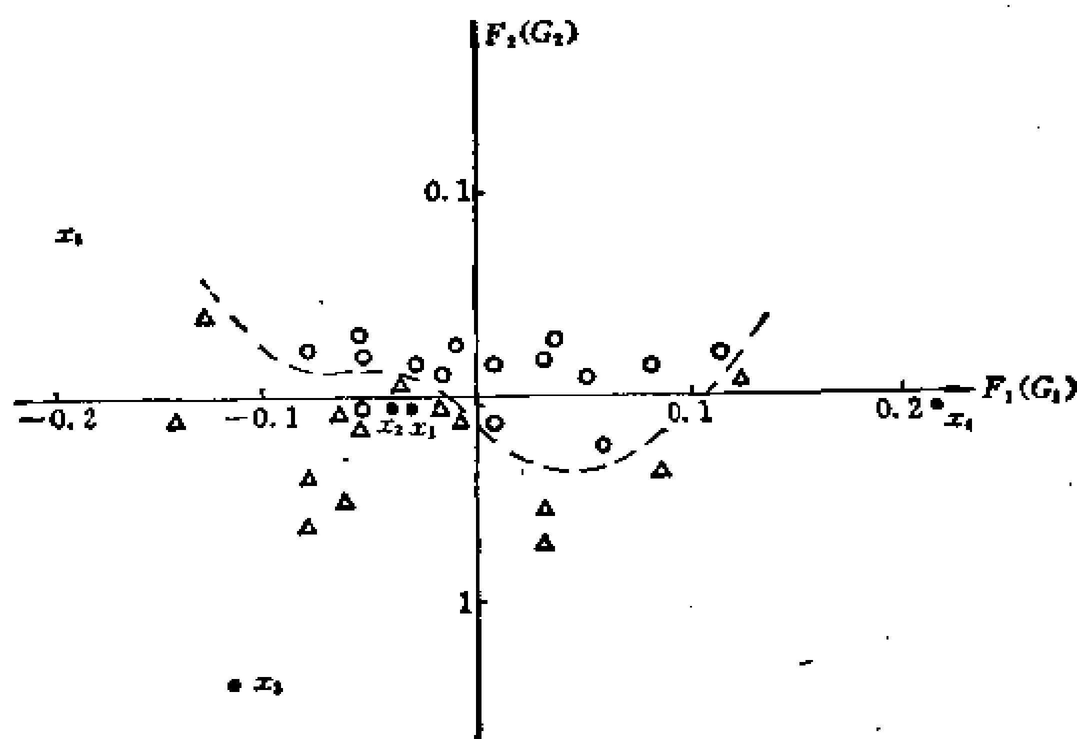


图 5.1　变量和样品在因子平面上的点聚图

表 5.3 Q 型因子载荷阵

样品号 \ 因子	G_1	G_2	样品号 \ 因子	G_1	G_2
1	−0.0100	0.0302	16	−0.0213	0.0120
2	0.0260	−0.0611	17	−0.0438	0.0391
3	0.0836	−0.0385	18	−0.0706	0.0244
4	−0.0844	−0.0285	19	−0.0874	−0.0545
5	−0.1258	0.0421	20	−0.0308	0.0057
6	−0.0429	0.0255	21	0.0288	0.0338
7	−0.0119	−0.0068	22	0.1122	0.0137
8	0.0772	0.0126	23	0.0023	−0.0123
9	−0.0573	−0.0346	24	0.0275	−0.0421
10	−0.0683	−0.0057	25	0.1190	0.0067
11	0.0428	0.0057	26	0.0537	−0.0294
12	−0.0159	0.0105	27	−0.0446	−0.0092
13	−0.0541	−0.0048	28	−0.0204	−0.0001
14	0.0224	0.0272	29	−0.1405	−0.0130
15	0.0416	0.0248			

(iv) 计算Q型的因子载荷．矩阵B头两个特征根仍为 $\lambda_1=0.1231$，$\lambda_2=0.0220$．相应特征向量为 $\mathbf{e}_1^*=Z'\mathbf{e}_1$，$\mathbf{e}_2^*=Z'\mathbf{e}_2$．然后再单位化．其载荷阵列在表 5.3．

(v) 将变量和样品根据它们的载荷同时点在以两个因子轴为坐标的平面上，如图 5.1 所示，图中样品记号△表示涝年，样品记号○表示旱年，变量 $x_1,\cdots,x_5$ 直接标出．

由图 5.1 可看出，变量 x_1,x_2 很接近表示它们有较密切的相关关系，而 x_3,x_4,x_5 它们之间相关关系较差．从样品点看，涝年相对集中在第三象限，此外在 x_1,x_2 附近的样品点特别多，表示了用 x_1 和 x_2 来解释六月份的旱涝关系比较合适．

习 题 十

10.1 设变量 x_1,x_2 和 x_3 已标准化，其相关系数矩阵为

$$R = \begin{bmatrix} 1.00 & & \\ 0.63 & 1.00 & \\ 0.45 & 0.35 & 1 \end{bmatrix}$$

试对变量进行正交因子分析。

10.2 在上题中取 $m=2$ 进行正交因子旋转。

10.3 试比较主成分分析和因子分析的相似之处与不同点。

10.4 证明定理 5.1。

*10.5 若你手头有计算机可使用，试对下列表中 20 个土壤样品，每个样品观测 5 个变量的数据进行对应分析。

20 个土壤样品数据表

样品号	含沙量 x_1	淤泥含量 x_2	粘土含量 x_3	有机物 x_4	pH 值 x_5
1	77.3	13.0	9.7	1.5	6.4
2	82.5	10.0	7.5	1.5	6.5
3	66.9	20.6	12.5	2.3	7.0
4	47.2	33.3	19.0	2.8	5.8
5	65.3	20.5	14.2	1.9	6.9
6	83.3	10.0	6.7	2.2	7.0
7	81.6	12.7	5.7	2.9	6.7
8	47.8	36.5	15.7	2.3	7.2
9	48.6	37.1	14.3	2.1	7.2
10	61.6	25.5	12.9	1.9	7.3
11	58.6	26.5	14.9	2.4	6.7
12	69.3	22.3	8.4	4.0	7.0
13	61.8	30.8	7.4	2.7	6.4
14	67.7	25.3	7.0	4.8	7.3
15	57.2	31.2	11.6	2.4	6.3
16	67.2	22.7	10.1	3.3	6.2
17	59.2	31.2	9.6	2.4	6.0
18	80.2	13.2	6.6	2.0	5.8
19	82.2	11.1	6.7	2.2	7.2
20	69.7	20.7	9.6	3.1	5.9

第十一章　典型相关分析

§11.1　概　述

我们知道，对于两个随机变量的相关关系的程度可用它们的相关系数来衡量，如果有显著的相关关系进而还可建立它们的线性回归方程．当考虑一个随机变量与一组随机变量之间的相关时，可用复相关系数来衡量．在第六章中我们讨论过多对多的线性回归方程，它实质上是研究一组随机变量与另一组随机变量之间的线性相关问题．本章是利用主成分的思想来讨论两组随机变量的相关性问题，即把两组变量之间的相关性研究化成少数几对变量之间的相关性研究，而且这少数几对变量之间又是不相关的，这样我们就可以把两组变量之间的复杂相关关系化简．这种思想是 Hotelling 1936 年发展起来的．更确切的说，我们要在第一组变量中找出一个变量的线性组合(综合变量)，在第二组变量中也找出一个变量的线性组合(也是综合变量)，使它们具有最大的相关，如果这一对综合变量还不能够反映两组变量之间的相关性，还可继续在每一组变量中找出第二个线性组合，使得在与第一个线性组合不相关的线性组合中具有最大的相关．将此程序继续下去，可以将两组变量间的相关提取完毕．当然在实际中，我们希望只提取少数几对就能足够反映两组变量之间的相关．

这种研究两组变量的相关性方法在实际中是非常有用的．例如在地质科学中，研究超基性岩时要了解铬尖晶石的化学成分(一组变量)与周围岩石(另一组变量)之间的关系，以考查铬尖晶石形成时的岩浆环境及它最终固定下来的围岩环境，从成因来说这种关系是十分重要的；又如生物科学中，要研究生物群（用一组变量刻画)与生活环境(另一组变量)之间的关系，这对于保持生态平衡

是有指导意义的．用典型相关分析也能对经济现象进行研究和分析(见例 3.3)．

§11.2 典型变量与典型相关

(一) 总体的典型变量与典型相关

考虑 p 维随机向量 $x=(x_1, x_2, \cdots, x_p)'$，其协方差矩阵为 $\Sigma(\Sigma>0)$，将 x 剖分为

$$x=\begin{bmatrix} x^{(1)} \\ x^{(2)} \end{bmatrix} \tag{2.1}$$

其中子向量 $x^{(1)}$, $x^{(2)}$ 分别为 p_1 维和 p_2 维，$p_1+p_2=p$，且设 $p_1 \leqslant p_2$．相应地将 Σ 剖分为

$$\Sigma=\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{2.2}$$

其中 Σ_{11} 和 Σ_{22} 分别为 $x^{(1)}$ 和 $x^{(2)}$ 的协方差阵，$\Sigma_{12}=\Sigma_{21}'$ 为 $x^{(1)}$ 与 $x^{(2)}$之间的协方差阵，由于 $\Sigma>0$，所以 $\Sigma_{ii}>0(i=1, 2)$．现在我们来研究两组随机变量 $x^{(1)}$ 与 $x^{(2)}$ 的典型相关问题．

考虑两组随机变量的线性组合

$$\left.\begin{aligned} u &= l'x^{(1)} = l_1x_1+\cdots+l_{p_1}x_{p_1} \\ v &= m'x^{(2)} = m_1x_{p_1+1}+\cdots m_{p_2}x_{p} \end{aligned}\right\} \tag{2.3}$$

其中 l 及 m 为任意的非零常数向量，其维数分别为 p_1 维和 p_2 维．我们要寻找 l 及 m 使得 u 与 v 之间的相关系数达到最大．因为两个随机变量乘以常数后并不改变其相关系数，不妨限定 u 与 v 的方差为 1．即

$$\left.\begin{aligned} \mathrm{var}(u) &= l'\mathrm{cov}(x^{(1)})l = l'\Sigma_{11}l = 1 \\ \mathrm{var}(v) &= m'\mathrm{cov}(x^{(2)})m = m'\Sigma_{22}m = 1 \end{aligned}\right\} \tag{2.4}$$

在此条件之下，u 与 v 的相关系数为

$$\rho_{u,v}=\mathrm{cov}(u,v)=l'\mathrm{cov}(x^{(1)},x^{(2)})m=l'\Sigma_{12}m \tag{2.5}$$

于是我们的问题就变成了在条件 (2.4) 之下寻找 l 和 m 使得 (2.5)达到最大(注意到，若 u、v 有最大负相关，只须取 $l^*=-l$

代替 $\boldsymbol{l}$ 后，$-u$、v 有最大正相关，今后只考虑最大正相关）利用 Lagrange 乘数法，令

$$Q = \boldsymbol{l}'\Sigma_{12}\boldsymbol{m} - \frac{1}{2}\lambda_1(\boldsymbol{l}'\Sigma_{11}\boldsymbol{l} - 1) - \frac{1}{2}\lambda_2(\boldsymbol{m}'\Sigma_{22}\boldsymbol{m} - 1)$$

其中 λ_1 与 λ_2 为 Lagrange 乘数，利用第一章第 4 节矩阵求导的知识，求 Q 关于 $\boldsymbol{l}$ 和 $\boldsymbol{m}$ 的导数，并令其为 0。

$$\left.\begin{aligned} \frac{\partial Q}{\partial \boldsymbol{l}} &= \Sigma_{12}\boldsymbol{m} - \lambda_1\Sigma_{11}\boldsymbol{l} = 0 \\ \frac{\partial Q}{\partial \boldsymbol{m}} &= \Sigma_{21}\boldsymbol{l} - \lambda_2\Sigma_{22}\boldsymbol{m} = 0 \end{aligned}\right\} \tag{2.6}$$

将(2.6)中的两式分别左乘 $\boldsymbol{l}'$ 和 $\boldsymbol{m}'$，并注意到条件(2.4)便可得

$$\left.\begin{aligned} \boldsymbol{l}'\Sigma_{12}\boldsymbol{m} &= \lambda_1\boldsymbol{l}'\Sigma_{11}\boldsymbol{l} = \lambda_1 \\ \boldsymbol{m}'\Sigma_{21}\boldsymbol{l} &= \lambda_2\boldsymbol{m}'\Sigma_{22}\boldsymbol{m} = \lambda_2 \end{aligned}\right\}$$

但 $(\boldsymbol{l}'\Sigma_{12}\boldsymbol{m})' = \boldsymbol{m}'\Sigma_{21}\boldsymbol{l}$，即 $\lambda_1' = \lambda_2$，而 λ_1 与 λ_2 都是实数。因此令 $\lambda = \lambda_1 = \lambda_2$，从而得到

$$\rho_{u,v} = \boldsymbol{l}'\Sigma_{12}\boldsymbol{m} = \boldsymbol{m}'\Sigma_{21}\boldsymbol{l} = \lambda$$

这说明了 Lagrange 乘数 λ_1 与 λ_2 相等，且等于 u 与 v 的相关系数。(2.6)式可写为

$$\left.\begin{aligned} \Sigma_{12}\boldsymbol{m} - \lambda\Sigma_{11}\boldsymbol{l} = 0 \\ \Sigma_{21}\boldsymbol{l} - \lambda\Sigma_{22}\boldsymbol{m} = 0 \end{aligned}\right\} \tag{2.7}$$

将(2.7)式中的第一式左乘 $\Sigma_{21}\Sigma_{11}^{-1}$ 得 $\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\boldsymbol{m} - \lambda\Sigma_{21}\boldsymbol{l} = 0$ 并把第二式代入得

$$\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\boldsymbol{m} - \lambda^2\Sigma_{22}\boldsymbol{m} = 0 \tag{2.8}$$

用 Σ_{22}^{-1} 左乘(2.8)，得

$$(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\boldsymbol{m} = \lambda^2\boldsymbol{m} \tag{2.9}$$

仿此，用 $\Sigma_{12}\Sigma_{22}^{-1}$ 左乘(2.7)中的第二式，并将第一式代入得

$$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\boldsymbol{l} - \lambda^2\Sigma_{11}\boldsymbol{l} = 0 \tag{2.10}$$

用 Σ_{11}^{-1} 左乘(2.10)得

$$(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\boldsymbol{l} = \lambda^2\boldsymbol{l} \tag{2.11}$$

记 $A = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, $B = \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, 那么(2.9)及(2.11)可写为

$$Bm = \lambda^2 m \tag{2.12}$$

$$Al = \lambda^2 l \tag{2.13}$$

显然，方程(2.6)化成了求A与B的特征根 λ^2 与特征向量问题。关于A与B的特征根和特征向量有如下定理：

定理 2.1 A与B有相同非零特征根，且特征根非负。

证明：记 $\Sigma_{11}^{-1}\Sigma_{12} = C_1$, $\Sigma_{22}^{-1}\Sigma_{21} = C_2$, 显然 C_1 为 $p_1 \times p_2$ 阶，C_2 为 $p_2 \times p_1$ 阶。由第一章推论 2.2 知道当 $\lambda \neq 0$ 时，

$$|\lambda I_{p_1} - C_1C_2| = \lambda^{p_1-p_2}|\lambda I_{p_2} - C_2 \cdot C_1|$$

注意到 $A = C_1C_2$, $B = C_2 \cdot C_1$. 因此A与B有相同的（包括重根）非零特征根。

又因为 $A = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ 与$G_1 = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$ 有相同的非零特征根，而后者是非负定阵，因此它们的特征根非负。 ▮

用 $\lambda_1^2 \geqslant \lambda_2^2 \geqslant \cdots \geqslant \lambda_s^2 > 0$ 表示A和B的非零特征根，$s = \mathrm{rk}A = \mathrm{rk}B$. 由于$A$与 $G_1 = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$ 有相同非零特征根，B与 $G_2 = \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$ 有相同的非零特征根，从而 G_1 与 G_2的非零特征根都是 $\lambda_1^2, \cdots, \lambda_s^2$. 令 $H = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$, 则$G$与$G_2$可表为

$$G_1 = HH', \quad G_2 = H'H \tag{2.14}$$

设 $a_1, a_2, \cdots, a_s$ 为 G_1 对应于非零 $\lambda_1^2, \lambda_2^2, \cdots, \lambda_s^2$ 的单位正交化特征向量，$b_1, b_2, \cdots, b_s$ 为 G_2 对应于 $\lambda_1^2, \lambda_2^2, \cdots, \lambda_s^2$ 的单位正交化特征向量。

定理 2.2 记号同上所述，令

$$l_i = \Sigma_{11}^{-\frac{1}{2}}a_i, m_i = \Sigma_{22}^{-\frac{1}{2}}b_i \ (i = 1, 2, \cdots, s) \tag{2.15}$$

则有

(i) l_i 是A对应于 λ_i^2 的特征向量，m_i 是B对应于 λ_i^2 的特征向量；

(ii) $l_i'\Sigma_{11}l_j = \begin{cases}1 & i = j \\ 0 & i \neq j\end{cases}$, $m_i'\Sigma_{22}m_j = \begin{cases}1 & i = j \\ 0 & i \neq j\end{cases}$ (2.16)

$$l_i'\Sigma_{12}m_j = \begin{cases}\lambda_i & i = j\\ 0 & i \neq j\end{cases} \tag{2.17}$$

其中 λ_i 是 λ_i^2 的正平方根.

证明: (i) 由于 a_i 为 G_1 对应于 λ_i^2 的特征向量,有 $G_1a_i = \lambda_i^2a_i$, 再注意到(2.15)即可得

$$(\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}})\Sigma_{11}^{\frac{1}{2}}l_i = \lambda_i^2\,\Sigma_{11}^{\frac{1}{2}}l_i$$

在上式两边左乘 $\Sigma_{11}^{-\frac{1}{2}}$, 得

$$(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})l_i = \lambda_i^2l_i$$

这说明了 l_i 是A对应于 λ_i^2 的特征向量. 同样可证明 m_i 是B对应于 λ_i^2 的特征向量.

(ii) 注意到(2.15)式以及 a_i 为单位化特征向量,所以有

$$a_i'a_j = l_i'\Sigma_{11}^{\frac{1}{2}}\Sigma_{11}^{\frac{1}{2}}l_j = l_i\Sigma_{11}l_j = \begin{cases}1 & i \neq j\\ 0 & i \neq j\end{cases}$$

同样道理(2.16)中的第二式也成立.

因为 a_i 是 $G_1 = HH'$ 的特征向量,所以有

$$HH'a_i = \lambda_i^2a_i \tag{2.18}$$

上式两边左乘 H' 得

$$(H'H)(H'a_i) = \lambda_i^2(H'a_i)$$

上式说明 $H'a_i$ 是 G_2 对应于 λ_i^2 的特征向量,而我们已设 G_2 对应于 λ_i^2 的单位化特征向量为 b_i, 由于 $a_i'a_i = b_i'b_i$, 所以有 $b_i = H'a_i/\lambda_i$, 再注意到(2.18)式就有

$$\begin{aligned} l_i'\Sigma_{12}m_j &= a_i'\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}b_j = a_i'HH'a_j/\lambda_j \\ &= \lambda_ja_i'a_j = \begin{cases}\lambda_i & i = j\\ 0 & i \neq j\end{cases}\end{aligned}$$

这便证明了(2.17)式 ▌

如果 $l = l_1$(p_1 维向量), $m = m_1$(p_2 维向量), 使得在条件(2.4)之下极大化 $\sigma_{u,v} = l'\Sigma_{12}m$ 那么

$$u_1 = l_1'x^{(1)},\ v_1 = m_1'x^{(2)} \tag{2.19}$$

称为总体的第一对典型变量,而

$$\rho_{u_1,v_1} = l_1'\Sigma_{12}m_1 \tag{2.20}$$

称为第一对典型变量的典型相关系数. 如果第一对典型变量还不能反映两组变量之间的相关性,还要找出第二对

$$u_2 = l_2'x^{(1)}, \quad v_2 = m_2'x^{(2)} \tag{2.21}$$

其中 l_2 和 m_2 除了满足条件(2.4)外,还要满足

$$\mathrm{cov}(l_1'x^{(1)},\ l_2'x^{(1)}) = l_1'\Sigma_{11}l_2 = 0$$

$$\mathrm{cov}(m_1'x^{(2)},\ m_2'x^{(2)}) = m_1'\Sigma_{22}m_2 = 0$$

且使得 $\rho_{u,v} = l'\Sigma_{12}m$ 达到极大. 称 u_2, v_2 为总体的第二对典型变量, $\rho_{u_2,v_2} = l_2'\Sigma_{12}m_2$ 为第二对典型变量的典型相关系数. 如此继续下去,第 $i(1 \leqslant i \leqslant s)$ 对典型相关变量为

$$u_i = l_i'x^{(1)}, \quad v_i = m_i'x^{(2)} \tag{2.22}$$

其中 l_i 和 m_i 除了满足条件(2.4)外,还要满足

$$\mathrm{cov}(l_i'x^{(1)},\ l_j'x^{(1)}) = 0 \quad (j = 1,2,\cdots,i-1)$$

$$\mathrm{cov}(m_i'x^{(2)},\ m_j'x^{(2)}) = 0 \quad (j = 1,2,\cdots,i-1)$$

且使得 $\rho_{u,v} = l'\Sigma_{12}m$ 达到极大. 称 u_i, $v_i(1 \leqslant i \leqslant p_1)$ 为总体的第 i 对典型变量, $\rho_{u_i,v_i} = l_i'\Sigma_{12}m_i$ 称为第 i 对典型变量的典型相关系数.

由上述总体典型变量和典型相关系数的定义以及前面的推导可知,在约束(2.4)之下求$l'\Sigma_{12}m$的极大,其极大值必为方程(2.13)的第一个大的特征根 λ_1^2 的正平方根 λ_1, 由定理2.2中的(2.17)式知 $l_1'\Sigma_{12}m_1 = \lambda_1$, 即极大值在 $l = l_1 = \Sigma_{11}^{-\frac{1}{2}}a_1$, $m = m_1 = \Sigma_{22}^{-\frac{1}{2}}b_1$ 时达到,再由定理 2.2 的 (i) 知 l_1 和 m_1 满足方程 (2.13) 和 (2.12). 从而我们获得第一对典型变量 $u_1 = l_1'x^{(1)} = (\Sigma_{11}^{-\frac{1}{2}}a_1)'x^{(1)}$, $v_1 = m_1'x^{(2)} = (\Sigma_{22}^{-\frac{1}{2}}b_1)'x^{(2)}$, 其中 a_1 与 b_1 分别为 G_1 和 G_2 对应于特征根 λ_1^2 的单位正交化特征向量典型相关系数为 $\rho_{u_1,v_1} = \lambda_1$. 类似地,第 $i(1 \leqslant i \leqslant s)$ 对典型变量为 $u_i = l_i'x^{(1)} = (\Sigma_{11}^{-\frac{1}{2}}a_i)'x^{(1)}$, $v_i = m_i'x^{(2)} = (\Sigma_{22}^{-\frac{1}{2}}b_i)'x^{(2)}$, 其典型相关系数为 $\rho_{u_i,v_i} = \lambda_i$. 将上面的讨论归纳为如下定理:

定理 2.3 设 $x^{(1)}$ 与 $x^{(2)}$ 分别为 p_1 维和 $p_2(p_1 \leqslant p_2)$ 维随

机向量，二阶矩存在，则它们的第 $i(1\leqslant i\leqslant s)$ 对典型变量为 $u_i=l_i'x^{(1)}$，$v_i=m_i'x^{(2)}$，其中 l_i 和 m_i 为(2.15)式所定义。它们的典型相关系数为 $\rho_{u_i,v_i}=\lambda_i$，其中 λ_i 为 A 的第 i 个大的特征根的正平方根。

在实际计算时，往往直接求 A 的特征根 λ_i^2 及相应的特征向量 l_i，并要求 $l_i'\Sigma_{11}l_i=1$，若不等于1，而等于 c_i，则只须用 $c_i^{-\frac{1}{2}}$ 乘 l_i 即可。根据(2.7)式，m_i 的计算可由 l_i 求得，即用如下公式求 M_i：

$$m_i=\frac{1}{\lambda_i}\Sigma_{22}^{-1}\Sigma_{21}l_i \tag{2.23}$$

同样要求 $m_i'\Sigma_{22}m_i=1$，否则类似 l_i 进行处理即可。

我们也可以从相关系数矩阵出发导出典型变量和典型相关系数，将相关矩阵 R 作相应的剖分：

$$R=\begin{bmatrix}R_{11} & R_{12}\\ R_{21} & R_{22}\end{bmatrix}$$

并记

$$D_1=\mathrm{diag}(\sqrt{\sigma_{11}},\cdots,\sqrt{\sigma_{p_1p_1}}),$$
$$D_2=\mathrm{diag}(\sqrt{\sigma_{p_1+1,p_1+1}},\cdots,\sqrt{\sigma_{pp}})$$

其中 σ_{ii} 为协方差阵 Σ 的对角元素，则有

$$\Sigma_{11}=D_1R_{11}D_1,\ \Sigma_{22}=D_2R_{22}D_2,\ \Sigma_{12}=D_1R_{12}D_2=\Sigma_{21}'$$

将它们代入 A 与 B 的表达式中便得到

$$A=\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}=D_1^{-1}R_{11}^{-1}R_{12}R_{22}^{-1}R_{21}D_1$$
$$B=\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}=D_2^{-1}R_{22}^{-1}R_{21}R_{11}^{-1}R_{12}D_2$$

记

$$A^*=R_{11}^{-1}R_{12}R_{22}^{-1}R_{21},\quad B^*=R_{22}^{-1}R_{21}R_{11}^{-1}R_{12} \tag{2.24}$$

从而有

$$A^*=D_1AD_1^{-1},\quad B^*=D_2BD_2^{-1} \tag{2.25}$$

故 A^* 与 A 有相同非零特征根，B^* 与 B 也有相同非零特征根，从而 A^* 与 B^* 有相同非零特征根，且为 $\lambda_1^2\geqslant\cdots\geqslant\lambda_s^2>0$，由

定理 2.2 的 (i) 有 $Al_i = \lambda_i^2 l_i$ $(i = 1, 2, \cdots, s)$，再由(2.25)有 $(D_1^{-1}A^*D_1)l_i = \lambda_i^2 l_i$，并在两边左乘 D_1 得

$$A^*(D_1 l_i) = \lambda_i^2(D_1 l_i)$$

上式说明 A^* 相应于特征根 λ_i^2 的特征向量为 $D_1 l_i$，记

$$l_i^* = D_1 l_i \tag{2.26}$$

同样，对于 B^* 相应于特征根 λ_i^2 的特征向量为

$$m_i^* = D_2 m_i \tag{2.27}$$

类似于(2.23)式，m_i^* 可由 l_i^* 来表达，即

$$m_i^* = \frac{1}{\lambda_i} R_{22}^{-1} R_{21} l_i^* \tag{2.28}$$

因此，由 A^* 求出它的特征根 λ_i^2 及相应的特征向量 l_i^*，再由(2.28)式又可求出 B^* 对应于 λ_i^2 的特征向量 m_i^*，然后根据(2.26)和(2.27)式就可写出变量组 $x^{(1)}$ 与 $x^{(2)}$ 的第 i 对典型变量

$$u_i = l_i' x^{(1)} = (D_1^{-1} l_i^*)' x^{(1)}, v_i = m_i' x^{(2)} = (D_2^{-1} m_i^*)' x^{(2)}$$
$$(i = 1, 2, \cdots s) \tag{2.29}$$

这里 $l_i = D_1^{-1} l_i^*$ 和 $m_i = D_2^{-1} m_i^*$ 仍要满足(2.4)，否则作适当处理。

（二） 典型变量和典型相关系数的估计

上一段的讨论是针对总体 $x = (x^{(1)'}, x^{(2)'})'$ 的协方差矩阵 Σ 已知的情况，当 Σ 未知时，就需要由样本进行估计。设 $x_i (i = 1, 2, \cdots, n)$ 是正态总体 $N_p(\mu, \Sigma)(\Sigma > 0)$ 的容量为 $n (n > p)$ 的样本，Σ 的极大似然估计为

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \tag{2.30}$$

其中

$$S_{11} = \frac{1}{n} \sum_{i=1}^{n} (x_i^{(1)} - \bar{x}^{(1)})(x_i^{(1)} - \bar{x}^{(1)})' \tag{2.31}$$

$$S_{22} = \frac{1}{n} \sum_{i=1}^{n} (x_i^{(2)} - \bar{x}^{(2)})(x_i^{(2)} - \bar{x}^{(2)})' \tag{2.32}$$

表 2.1　小麦 12 个变量的

	1	2	3	4	5
1	1.0000				
2	0.7025	1.0000			
3	0.0539	−0.0717	1.0000		
4	0.1154	0.0458	0.5811	1.0000	
5	−0.0669	−0.1212	0.5851	0.4065	1.0000
6	−0.0361	−0.0515	0.7137	0.5238	0.6698
7	0.4381	0.6109	−0.2064	−0.1113	−0.4702
8	0.1332	0.1667	−0.0708	−0.1186	−0.0686
9	0.4611	0.5927	−0.2545	−0.1213	−0.4649
10	0.5139	0.6633	−0.3099	−0.1602	−0.3441
11	0.4197	0.5148	−0.1491	−0.0216	−0.3475
12	0.6601	0.7129	−0.1652	−0.0121	−0.3632

表 2.2　六对样本典型变量的

序　号	典型相关系数 $\hat{\lambda}_i$	典型变量的	
1	0.8602	$\hat{l}_1=(0.3259,$	0.2433
		$\hat{m}_1=(0.2007,$	$-0.1091,$
2	0.6455	$\hat{l}_2=(-0.0918,$	$0.0206,$
		$\hat{m}_2=(0.1722,$	$-0.0149,$
3	0.5173	$\hat{l}_3=(-0.6383,$	$0.7164,$
		$\hat{m}_3=(0.0647,$	$0.4644,$
4	0.3078	$\hat{l}_4=(-0.0495,$	$0.1723,$
		$\hat{m}_4=(-0.1368,$	$0.3071,$
5	0.1727	$\hat{l}_5=(0.1976,$	$-0.1811,$
		$\hat{m}_5=(-0.6824,$	$-0.0100,$
6	0.0041	$\hat{l}_6=(0.1997,$	$-0.1200,$
		$\hat{m}_6=(0.5137,$	$-0.1753,$

相关系数矩阵

6	7	8	9	10	11	12
1.0000						
−0.2029	1.0000					
−0.1693	0.3503	1.0000				
−0.2284	0.8857	0.2945	1.0000			
−0.2141	0.8295	0.2899	0.9339	1.0000		
−0.1929	0.7007	0.0252	0.6960	0.6987	1.0000	
−0.1119	0.8155	0.0761	0.8686	0.8474	0.8015	1.0000

系数向量及典型相关系数

系数向量 $\hat{L}_i,\hat{M}_i$			
−0.2006,	−0.0225,	−0.1817,	0.2791)′
−0.4965,	0.3983,	−0.2696,	0.6856)′
0.2266,	0.0516,	−0.3612,	−0.0173)′
0.6693,	−0.6922,	0.1297,	−0.1621)′
0.1486,	−0.1055,	0.2198,	−0.4804)′
−0.3085,	0.5517,	0.3614,	−0.5000)′
0.3508,	0.1312,	0.1222,	−0.2238)′
−0.3921,	−0.4246,	0.1116,	0.7352)′
−0.0930,	0.2838,	0.1290,	−0.4349)′
0.5180,	−0.0377,	0.4874,	−0.1640)′
0.3284,	−0.3387,	0.0487,	−0.1920)′
−0.6777,	0.2756,	0.2852,	−0.2982)′

$$S_{12}=\frac{1}{n}\sum_{i=1}^{n}(x_i^{(1)}-\bar{x}^{(1)})(x_i^{(2)}-\bar{x}^{(2)})' \tag{2.33}$$

由第三章引理 2.1 知 S 以概率 1 正定，所以 S_{11}^{-1}, S_{22}^{-1} 存在，由极大似然估计的不变性知，由 S 产生的

$$\hat{A}=S_{11}^{-1}S_{12}S_{22}^{-1}S_{21},\quad \hat{B}=S_{22}^{-1}S_{21}S_{11}^{-1}S_{12} \tag{2.34}$$

分别是 A 和 B 的极大似然估计. 与总体协方差阵 Σ 已知时的情况相仿，我们可证明 $\hat{A}$ 及 $\hat{B}$ 有相同的非零特征根 $\hat{\lambda}_1^2\geqslant\cdots\geqslant\hat{\lambda}_{p_1}^2>0(p_1=\text{rk}\hat{A})$，且它们是 $\lambda_1^2,\cdots,\lambda_{p_1}^2$ 的极大似然估计，相应的特征向量 $\hat{\boldsymbol{l}}_1,\cdots,\hat{\boldsymbol{l}}_{p_1}$ 是 $\boldsymbol{l}_1,\cdots,\boldsymbol{l}_{p_1}$ 的极大似然估计，$\hat{\boldsymbol{m}}_1,\cdots,\hat{\boldsymbol{m}}_{p_1}$ 是 $\boldsymbol{m}_1,\cdots,\boldsymbol{m}_{p_1}$ 的极大似然估计. 于是可得到样本的典型相关变量:

$$\hat{u}_i=\hat{\boldsymbol{l}}_i'x^{(1)},\ \hat{v}_i=\hat{\boldsymbol{m}}_i x^{(2)}\ (i=1,2,\cdots,p_1) \tag{2.35}$$

类似于(2.23)式，$\hat{\boldsymbol{m}}_i$ 可由下式进行计算

$$\hat{\boldsymbol{m}}_i=\frac{1}{\hat{\lambda}_i}S_{22}^{-1}S_{21}\hat{\boldsymbol{l}} \tag{2.36}$$

以及相应的样本典型相关系数

$$\hat{\rho}_{\hat{u}_i,\hat{v}_i}=\hat{\lambda}_i\ (i=1,2,\cdots,p_1) \tag{2.37}$$

(当 $\text{rk}(A)=s<p_1$ 时我们只取前 s 个 $\hat{\lambda}_1,\cdots,\hat{\lambda}_s$ 作为非零特征根 $\lambda_1,\cdots,\lambda_s$ 的极大似然估计，特征向量也有类似结论).

当然，我们也可以从样本相关系数矩阵出发导出样本典型变量和样本典型相关系数. 设样本相关系数为

$$\hat{R}=\begin{bmatrix}\hat{R}_{11} & \hat{R}_{12}\\ \hat{R}_{21} & \hat{R}_{22}\end{bmatrix}$$

令 $S_1=\text{diag}(\sqrt{s_{11}},\cdots,\sqrt{s_{p_1p_1}})$, $S_2=\text{diag}(\sqrt{s_{p_1+1,p_1+1}},\cdots,\sqrt{s_{pp}})$ 其中 s_{ii} 为样本协方差阵 S 的对角元素，类似于总体的讨论可以得

$$\hat{A}^*=S_1\hat{A}S_1^{-1},\quad \hat{B}^*=S_2\hat{B}S_2^{-1} \tag{2.38}$$

其中

$$\hat{A}^*=\hat{R}_{11}^{-1}\hat{R}_{12}\hat{R}_{22}^{-1}\hat{R}_{21},\quad \hat{B}^*=\hat{R}_{22}^{-1}\hat{R}_{21}\hat{R}_{11}^{-1}\hat{R}_{12} \tag{2.39}$$

而 $\hat{A}$ 和 $\hat{B}$ 由(2.34)所定义. 故 $\hat{A}^*$ 与 $\hat{A}$ 有相同非零特征根，

$\hat{B}^*$ 与 $\hat{B}$ 也有相同非零特征根，从而 $\hat{A}^*$ 与 $\hat{B}^*$ 有相同非零特征根 $\hat{\lambda}_1^2,\cdots,\hat{\lambda}_p^2$，并可导出 $\hat{A}^*$ 对应于 $\hat{\lambda}_i^2$ 的特征向量 $\hat{l}_i^*$ 为

$$\hat{l}_i^* = S_1 \hat{l}_i \tag{2.40}$$

$\hat{B}^*$ 对应于 $\hat{\lambda}_i^2$ 的特征向量 $\hat{M}_i^*$ 为

$$\hat{m}_i^* = S_2 \hat{m}_i \tag{2.41}$$

类似于(2.28)有

$$\hat{m}_i^* = \frac{1}{\hat{\lambda}_i} \hat{R}_{22}^{-1} \hat{R}_{21} \hat{l}_i^* \tag{2.42}$$

于是第 $i(i=1,2,\cdots,p_1)$ 对样本典型变量为

$$\hat{u}_i = (S_1^{-1} \hat{l}_i^*)' x^{(1)},\ \hat{v}_i = (S_2^{-1} \hat{m}^*)' x^{(2)} \tag{2.43}$$

样本典型相关系数为 $\hat{\lambda}_i$。

例 2.1　N.Giri 给出了根据 27 株小麦样品 12 个指标的数据算出的样本相关系数矩阵列在表 2.1，考查前 6 个指标与后 6 个指标组成的两组变量之间的相关，所得的样本典型相关系数及样本典型变量的系数列在表 2.2。

§11.3　典型相关系数的检验及典型回归

（一）　典型相关系数的检验

我们的目的在于选出 m 对(一般 m 比 p_1 小得多) 典型变量来代替两组变量 $x^{(1)}$ 与 $x^{(2)}$ 的相关性的研究，而这 m 对典型变量的每对之间必须具有显著相关才行．我们知道第 i 对典型变量的典型相关系数为 $\hat{\lambda}_i$，那么 $\hat{\lambda}_i$ 为多大(或者 $1-\hat{\lambda}_i$ 为多小) 才算有显著相关，这需要进行检验．

设 $x=(x^{(1)\prime}x^{(2)\prime})'$ 服从正态 $N_p(\mu,\Sigma)$ $(\Sigma>0)$ 分布，又 $\lambda_1,\cdots,\lambda_{p_1}$ 为总体典型相关系数，检验假设

$$H_0: \lambda_1=\lambda_2=\cdots=\lambda_{p_1}=0,\ H_1:\ \lambda_i\ \text{不全为}\ 0 \tag{3.1}$$

等价于检验假设(见习题 11.4)

$$H_0: \Sigma_{12}=0,\ H_1: \Sigma_{12}\neq 0 \tag{3.2}$$

检验假设(3.2)就是检验子向量 $x^{(1)}$ 与 $x^{(2)}$ 的独立性问题，由第

五章第 5 节知检验统计量为

$$\Lambda_1^* = \frac{|A|}{|A_{11}||A_{22}|} \tag{3.3}$$

这里的 $A = nS = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'$, $A_{jj} = nS_{jj} = \sum_{i=1}^{n}(x_i^{(j)} - \bar{x}^{(j)})(x_i^{(j)} - \bar{x}^{(j)})'$ $(j = 1, 2)$.

检验统计量(3.3)等价于如下的统计量

$$\Lambda_1 = \frac{|S|}{|S_{11}||S_{22}|} \tag{3.4}$$

利用分块矩阵的行列式公式不难证明(习题 11.5)

$$|S| = |S_{22}||S_{11}|\prod_{i=1}^{p_1}(1 - \hat{\lambda}_i^2) \tag{3.5}$$

其中 $\hat{\lambda}_i^2 (i = 1, 2, \cdots, p_1)$ 为 $\hat{A} = S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}$ 的特征根，这样检验统计量(3.4)可写为

$$\Lambda_1 = \prod_{i=1}^{p_1}(1 - \hat{\lambda}_i^2) \tag{3.6}$$

1941 年 Bartlett 给出了在原假设 H_0 为真时 Λ_1 的一个渐近分布：

$$P(-\beta \ln \Lambda_1 \leqslant x) \sim P(\chi^2(f_1) \leqslant x) \tag{3.7}$$

其中$\beta = n - \frac{1}{2}(p_1 + p_2 + 1) = n - \frac{1}{2}(p + 1)$, $f_1 = p_1 \cdot p_2$

利用(3.7)对于显著水平 α，当 $-\beta \ln \Lambda_1 > \chi_{1-\alpha}^2(f)$ 时，拒绝 H_0 即至少第一对典型变量显著相关.

在假设 H_0 被拒绝后，Bartlett 进而考虑检验假设 H_1: $\lambda_2 = \lambda_3 = \cdots = \lambda_{p_1} = 0$，并使用检验统计量

$$\Lambda_2 = \prod_{i=2}^{p_1}(1 - \hat{\lambda}_i^2) \tag{3.8}$$

对于显著水平 α，根据是否满足

$$-\beta \ln \Lambda_2 > \chi^2_{1-\alpha}(f_2) \tag{3.9}$$

而决定是否拒绝 H_1. 其中 $f_2 = (p_1 - 1)\cdot(p_2 - 1)$, $\beta = n - \frac{1}{2}(p+1)$.

一般地，若前 $m-1$ 对典型变量是显著相关，进而要检验第 m 对是否显著相关，即检验假设 $H_0: \lambda_m = \lambda_{m+1} = \cdots = \lambda_{p_1} = 0$，使用的检验统计量为

$$\Lambda_m = \sum_{i=m}^{p_1}(1 - \hat{\lambda}_i^2) \tag{3.10}$$

对于显著水平 α，根据是否满足

$$-\beta \ln \Lambda_m > \chi^2_{1-\alpha}(f_m) \tag{3.11}$$

而决定是否拒绝 H_1. 其中 $f_m = (p_1 - m + 1)(p_2 - m + 1)$, $\beta = n - \frac{1}{2}(p+1)$.

例 3.1　对于例 2.1 所给出的六对典型变量，用 Bartlett 的方法进行显著性检验.

首先检验第一对典型变量，计算

$$-\beta \ln \Lambda_1 = -\left(27 - \frac{1}{2}(12+1)\right)\ln \prod_{i=1}^{6}(1 - \hat{\lambda}_i^2) = 47.7$$

对于 $\alpha = 0.1$, $\chi^2_{0.9}(36) = 47.212$, $-\beta \ln \Lambda_1 > \chi^2_{0.9}(36)$，所以拒绝 H_0，即第一对典型变量是显著相关，接着对第二对典型变量进行检验。计算

$$-\beta \ln \Lambda_2 = -\left(27 - \frac{1}{2}(12+1)\right)\ln \prod_{i=2}^{6}(1 - \hat{\lambda}_i^2) = 20.09$$

对于 $\alpha = 0.1$, $\chi^2_{0.9}(25) = 34.38$, $-\beta \ln \Lambda_2 < \chi^2_{0.9}(25)$，所以接受 H_0，即第二对典型变量不显著相关，从而其后的四对典型变量都不相关。

（二） 典型相关分析的步骤及例

为了方便实际计算，我们将典型相关分析的计算步骤归纳如下：

从样本协方差阵出发的计算步骤。

（1）由两组变量的资料阵

$$X=\begin{bmatrix}X^{(1)}\\X^{(2)}\end{bmatrix}=\begin{bmatrix}x_{11} & x_{12} & \cdots & x_{1n}\\ \cdots & \cdots & \cdots & \cdots\\ x_{p_1 1} & x_{p_1 2} & \cdots & x_{p_1 n}\\ \hline x_{p_1+1\,1} & x_{p_1+1\,2} & \cdots & x_{p_1+1\,n}\\ \cdots & \cdots & \cdots & \cdots\\ x_{p1} & x_{p2} & \cdots & x_{pn}\end{bmatrix},$$

$$(p_1\leqslant p_2=p-p_1)$$

按公式 $S_{ij}=\frac{1}{n}\sum_{\alpha=1}^{n}(x_{i\alpha}-\bar{x}_i)(x_{j\alpha}-\bar{x}_j)$，其中 $\bar{x}_i=\frac{1}{n}\sum_{\alpha=1}^{n}x_{i\alpha}$ 计算出样本协方差矩阵

$$S=\{S_{ij}\}_{p\times p}=\begin{bmatrix}S_{11} & S_{12}\\S_{21} & S_{22}\end{bmatrix}$$

（2）计算 $\hat{A}=S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}$ 的 p_1 个特征根：$\hat{\lambda}_1^2\geqslant\hat{\lambda}_2^2\geqslant\cdots\geqslant\hat{\lambda}_{p_1}^2$ 和相应的特征向量 $\boldsymbol{l}_1,\cdots,\boldsymbol{l}_{p_1}$，并要求 $\boldsymbol{l}_i'S_{11}\boldsymbol{l}_i=1$，若不等于 1 而等于 c_i，则将 $\boldsymbol{l}_i$ 乘以 $c_i^{-\frac{1}{2}}$ 即可。

（3）按公式(2.36)计算 $\hat{B}=S_{22}^{-1}S_{21}S_{11}^{-1}S_{12}$ 相对应于 $\hat{\lambda}_1^2,\cdots,\hat{\lambda}_{p_1}^2$ 的特征向量 $\hat{\boldsymbol{m}}_1,\cdots,\hat{\boldsymbol{m}}_{p_1}$，同样要求 $\hat{\boldsymbol{m}}_i'S_{22}\hat{\boldsymbol{m}}_i=1$，若不等于 1 而等于 d_i，则将 $\hat{\boldsymbol{m}}_i$ 乘以 $d_i^{-\frac{1}{2}}$。

（4）由(2)及(3)可得 p_1 对典型变量

$$\hat{u}_i=\boldsymbol{l}_i'\boldsymbol{x}^{(1)},\ \hat{v}_i=\hat{\boldsymbol{m}}_i\boldsymbol{x}^{(2)}\ (i=1,2,\cdots,p_1)$$

其相应的典型相关系数为 $\hat{\lambda}_i,(i=1,2,\cdots,p_1)$。

（5）利用 Bartlett 方法对典型变量是否显著相关进行检验。如果前 m 对典型变量显著相关，则取前 m 对典型变量代替原变量组 $\boldsymbol{x}^{(1)}$ 与 $\boldsymbol{x}^{(2)}$ 的研究。

(6) 根据具体问题对典型变量进行解释(参看后面的例子).

从样本相关系数矩阵出发的计算步骤.

(1) 计算样本相关系数矩阵

$$\hat{R}=\{\hat{r}_{ij}\}_{p\times p}=\begin{bmatrix}\hat{R}_{11} & \hat{R}_{12}\\ \hat{R}_{21} & \hat{R}_{22}\end{bmatrix}$$

其中 $\hat{r}_{ij}=s_{ij}/\sqrt{s_{ii}}\cdot\sqrt{s_{jj}}$, $s_{ij}=\frac{1}{n}\sum_{\alpha=1}^{n}(x_{i\alpha}-\bar{x}_i)(x_{j\alpha}-\bar{x}_j)$,$(i, j=1,2,\cdots,p)$

(2) 计算 $\hat{A}^*=\hat{R}_{11}^{-1}\hat{R}_{12}\hat{R}_{22}^{-1}\hat{R}_{21}$ 的 p_1 个特征根: $\hat{\lambda}_1^2\geqslant\hat{\lambda}_2^2\geqslant\cdots\geqslant\hat{\lambda}_{p_1}^2$ 与相应的特征向量 $\hat{l}_i^*(i=1,2,\cdots,p_1)$. 并要求 $\hat{l}_i^{*\prime}\hat{R}_{11}\hat{l}_i^*=1$, 若不等于1, 而等于 c_i 则用 $c_i^{-\frac{1}{2}}$ 去乘 $\hat{l}_i^*$ 即可. 令 $S_1=\mathrm{diag}(\sqrt{s_{11}},\cdots\sqrt{s_{p_1p_1}})$,再计算

$$\hat{l}_i=S_1^{-1}\hat{l}_i^*\quad i=1,2,\cdots,p_1$$

(3) 按公式(2.41)计算 $\hat{B}^*=\hat{R}_{22}^{-1}\hat{R}_{21}\hat{R}_{11}^{-1}\hat{R}_{12}$ 对应于 $\hat{\lambda}_1^2,\cdots,\hat{\lambda}_{p_1}^2$ 的特征向量 $\hat{m}_i^*(i=1,2,\cdots p_1)$, 要求 $\hat{m}_i^{*\prime}\hat{R}_{22}\hat{m}_i^*=1$, 否则,类似(2)进行处理. 令 $S_2=\mathrm{diag}(\sqrt{s_{p_1+1,p_1+1}},\cdots,\sqrt{s_{pp}})$. 计算

$$\hat{m}_i=S_2^{-1}\hat{m}_i^*\quad i=1,2,\cdots,p_1$$

(4) 由(2)和(3)便得到 p_1 对典型变量

$$\hat{u}_i=\hat{l}_i'x^{(1)},\quad \hat{v}_i=\hat{m}_i'x^{(2)}\quad(i=1,2,\cdots,p_1)$$

并用 Bartlett 方法进行检验.

(5) 对典型变量进行解释.

例 3.2 为了解某矿区下部的 $x_1=\mathrm{Pt}$, $x_2=\mathrm{Pd}$ 与 $x_3=\mathrm{Cu}$, $x_4=\mathrm{Ni}$ 的共生组合规律, 从钻孔中取出 27 个样品,测试了这 4 个变量的含量,利用典型相关分析研究 $(x_1,x_2)'$ 与 $(x_3,x_4)'$ 的相关关系.

(1) 由 27 个样品所测得的数据计算得 $(x_1,x_2,x_3,x_4)'$ 的相关系数矩阵为

$$\hat{R}=\left[\begin{array}{cc|cc}1.0000 & 0.6209 & 0.7471 & 0.8691\\ 0.6209 & 1.0000 & 0.4477 & 0.5273\\ \hline 0.7471 & 0.4477 & 1.0000 & 0.6818\\ 0.8691 & 0.5273 & 0.6818 & 1.0000\end{array}\right]=\begin{bmatrix}\hat{R}_{11} & \hat{R}_{12}\\ \hat{R}_{21} & \hat{R}_{22}\end{bmatrix}$$

(2) 计算

$$\hat{A}^*=\hat{R}_{11}^{-1}\hat{R}_{12}\hat{R}_{22}^{-1}\hat{R}_{21}=\begin{bmatrix}0.8130 & 0.4916\\ 0.0210 & 0.0126\end{bmatrix}$$

以及它的特征根 $\hat{\lambda}_1^2=(0.8946)^2$，$\hat{\lambda}_2^2=(0.0095)^2$，再计算其相应的特征向量 $\hat{l}_1^*=(1.0161,-0.0263)'$，$\hat{l}_2^*=(-0.7712,1.2754)'$ 它们满足 $\hat{l}_i^{*\prime}\hat{R}_{11}\hat{l}_i^*=1$ $(i=1,2)$。又 $S_1=\text{diag}(0.3890,0.2444)$ $S_2=\text{diag}(0.1216,\ 0.0723)$ 从而得 $l_1=S_1^{-1}\hat{l}_1^*=(0.9992,-0.0412)'$，$l_2=S_1^{-1}\hat{l}_2^*=(-0.3551,0.9348)'$。

(3) 按公式(2.41)可计算得 $\hat{B}^*=\hat{R}_{22}^{-1}\hat{R}_{21}\hat{R}_{11}^{-1}\hat{R}_{12}$ 相应于特征根 $\lambda_1^2=(0.8946)^2$，$\lambda_2^2=(0.0095)^2$ 的特征向量为

$$\hat{m}_1^*=\frac{1}{\lambda_1}R_{22}^{-1}R_{21}\hat{l}_1^*=(0.3232,\ 0.7514)'$$

$$\hat{m}_2^*=\frac{1}{\lambda_2}R_{22}^{-1}R_{21}\hat{l}_2^*=(-1.3281,\ 1.1418)'$$

它们满足 $\hat{m}_i^{*\prime}\hat{R}_{22}\hat{m}_i^*=1$ $(i=1,2)$。从而得

$$\hat{m}_1=S_2^{-1}\hat{m}_1^*=(0.2478,\ 0.9688)'$$
$$\hat{m}_2=S_2^{-1}\hat{m}_2^*=(-0.5688,0.8225)'$$

(4) 由(2)及(3)我们得到了两对典型变量及相应的典型相关系数：

$$\hat{u}_1=\hat{l}_1'x^{(1)}=0.9992x_1-0.0412x_2$$
$$\hat{v}_1=\hat{m}_1'x^{(2)}=0.2478x_3+0.9688x_4$$
$$\rho_{\hat{u}_1,\hat{v}_1}=\hat{\lambda}_1=0.8946$$
$$\hat{u}_2=\hat{l}_2'x^{(1)}=-0.3551x_1+0.9348x_2$$
$$\hat{v}_2=\hat{m}_2'x^{(2)}=-0.5688x_3+0.8225x_4$$
$$\rho_{\hat{u}_2,\hat{v}_2}=\hat{\lambda}_2=0.0095$$

（5）典型相关系数的显著性检验

$$\Lambda_1 = \prod_{i=1}^{2}(1-\hat{\lambda}_i^2) = (1-0.8003)(1-0.00009) = 0.1997$$

$$-\left[n-\frac{1}{2}(p+1)\right]\ln \Lambda_1 = -[27-2.5]\cdot(-1.6109)$$

$$= 39.4671$$

查自由度为 $p_1 \times p_2 = 4$，$\alpha = 0.05$ 的 χ^2 分布表得 $\chi^2_{0.95}(4) = 9.488$ 显然 $39.4671 > \chi^2_{0.95}(4)$。故第一对典型变量显著相关，利用(3.11)式类似可检验第二对典型变量，检验结果是不显著。因此只取第一对典型变量来代替原变量组的研究。

（6）解释

由于第一对典型变量显著相关，说明该矿下部的 Pt，Pd 与 Cu、Ni 是密切相关的，而典型变量的意义主要由那些系数较高的变量决定，$\hat{u}_1 = 0.9992x_1 - 0.0412x_2$ 中 x_1(Pt) 的系数为 0.9992 相对于 x_2(Pd) 的系数大得多，而 $\hat{v}_1 = 0.2478x_3 + 0.9688x_4$ 中，x_4(Ni) 的系数 0.9688 相对于 x_3(Cu) 的系数大，因此两组变量 (Pt，Pd) 与 (Cu，Ni) 的相关主要是由 Pt 与 Ni 的相关所体现，该矿的 Pt、Pd 的含量受着 Ni 矿化的影响，主要赋存于 Ni 的硫化物中，故以硫镍铂矿的形式出现。

例 3.3　研究 1953～1982 年共 30 年的投资性变量与国民收入变量之间的相关关系，投资性变量取五个：x_1 = 工业劳动者人数(万人)，x_2 = 农业劳动者人数(万人)，x_3 = 货物周转量(亿吨公里)，x_4 = 生产性建设投资（亿元），x_5 = 建筑安装工程投资(亿元)；国民收入变量取四个：x_6 = 农业国民收入(亿元)，x_7 = 工业国民收入(亿元)，x_8 = 建筑业国民收入(亿元)，x_9 = 运输业国民收入(亿元)．设第一组变量 $x^{(1)} = (x_1, x_2, x_3, x_4, x_5)'$，第二组变量 $x^{(2)} = (x_6, x_7, x_8, x_9)'$．根据 30 年的数据计算样本相关系数矩阵：

$$\hat{R}_{11}=\begin{pmatrix}1.0000 & & & & \\ 0.6905 & 1.0000 & & & \\ 0.9305 & 0.8231 & 1.0000 & & \\ 0.8530 & 0.5900 & 0.7875 & 1.0000 & \\ 0.9331 & 0.6629 & 0.9477 & 0.8848 & 1.0000\end{pmatrix}$$

$$\hat{R}_{22}=\begin{pmatrix}1.0000 & & & \\ 0.9526 & 1.0000 & & \\ 0.9285 & 0.9699 & 1.0000 & \\ 0.8648 & 0.9693 & 0.9538 & 1.0000\end{pmatrix}$$

$$\hat{R}_{21}=\hat{R}'_{12}=\begin{pmatrix}0.8528 & 0.8640 & 0.9547 & 0.6257 & 0.8600\\ 0.9359 & 0.8424 & 0.9963 & 0.8127 & 0.9500\\ 0.9419 & 0.7535 & 0.9639 & 0.8150 & 0.9637\\ 0.9455 & 0.7589 & 0.9606 & 0.9094 & 0.9676\end{pmatrix}$$

此时由于 $p_2=4$，$p_1=5$，即 $p_2<p_1$，故先计算 $\hat{B}^*=\hat{R}_{22}^{-1}\hat{R}_{21}\cdot\hat{R}_{11}^{-1}\hat{R}_{12}$ 的特征根及特征向量，再计算 $\hat{A}^*$ 相应的特征向量。最后得到 4 对典型变量及相应的典型相关系数：

$$\hat{u}_1=0.0931x_1+0.2025x_2+0.5511x_3-0.0729x_4+0.2640x_5$$

$$\hat{v}_1=0.1105x_6+0.9423x_7-0.0349x_8-0.0149x_9$$

$$\rho_{\hat{u}_1\hat{v}_1}=\hat{\lambda}_1=0.999$$

$$\hat{u}_2=-0.4918x_1-1.1585x_2+1.5426x_3+2.1708x_4-1.9262x_5$$

$$\hat{v}_2=-2.1973x_6+0.2742x_7+0.3203x_8+1.5936x_9$$

$$\rho_{\hat{u}_2\hat{v}_2}=\hat{\lambda}_2=0.965$$

$$\hat{u}_3=1.7784x_1+1.0352x_2-5.8838x_3-2.1792x_4+5.3670x_5$$

$$\hat{v}_3=0.5439x_6-4.5367x_7+3.9878x_8+0.1579x_9$$

$$\rho_{\hat{u}_3\hat{v}_3}=\hat{\lambda}_3=0.738$$

$$\hat{u}_4=-0.0995x_1+0.3392x_2-0.0674x_3+0.1753x_4-0.0353x_5$$

$\hat{v}_4 = 0.0645x_6 - 0.0068x_7 - 0.0203x_8 + 0.0942x_9$

$\rho_{\hat{u}_4\hat{v}_4} = \hat{\lambda}_4 = 0.129$

按 Bartlett 方法进行检验典型相关系数的显著性检验，在 $\alpha = 0.05$ 之下，只有前三对典型变量显著相关。因此取前三对典型变量进行研究，从前三对典型相关变量的系数可以看出，第一对典型变量主要反映了货物周转量 x_3 与工业国民收入 x_7 有较高的显著相关；第二对典型变量主要反映了生产性投资 x_4 与运输业国民收入 x_9 有较显著的相关性；第三对典型变量主要反映建筑安装工程投资 x_5 与建筑业国民收入 x_8 有显著的相关。

（三） 典型变量的回归

当我们获得了前 m 对典型相关变量 $\hat{u}_i, \hat{v}_i$ $(i = 1, 2, \cdots, m)$，且每对都具有显著的相关。这样我们进一步还可建立 $\hat{u}_i$ 与 $\hat{v}_i$ 之间的线性回归方程，这就是典型变量的回归。这种回归方程在地质学中的趋势分析，在生态学的系统分析中都有应用。因为 $\hat{u}_1$, $\hat{v}_1$ 的典型相关系数 $\hat{\lambda}_1$ 是最大，我们以 $\hat{u}_1, \hat{v}_1$ 为例来说明如何建立 $\hat{u}_1$ 与 $\hat{v}_1$ 的线性回归方程。

设 $x_i = (x_i^{(1)'}, x_i^{(2)'})'$ $i = 1, 2, \cdots, n$ 是容量为 n 的样本，并根据上一节的方法得到了第一对典型变量：

$$\hat{u}_1 = \hat{l}'x^{(1)}, \quad \hat{v}_1 = \hat{m}'x^{(2)} \tag{3.12}$$

将 n 个样品点分别代入 $\hat{u}_1$ 和 $\hat{v}_1$，得到 n 个样品的典型变量得分

$$(\hat{u}_{1j} = \hat{l}'x_j^{(1)}, \quad \hat{v}_{1j} = \hat{m}_1'x_j^{(2)}) \quad j = 1, 2, \cdots, n \tag{3.13}$$

由于 $\hat{u}_1$ 与 $\hat{v}_1$ 具有显著相关，故可建立回归方程

$$\hat{u}_1 = \hat{\beta}_0 + \hat{\beta}_1\hat{v}_1 \tag{3.14}$$

其中

$$\hat{\beta}_1 = \frac{n\sum_{j=1}^{n}(\hat{u}_{1j}\hat{v}_{1j}) - \left(\sum_{j=1}^{n}\hat{u}_{1j}\right)\left(\sum_{j=1}^{n}\hat{v}_{1j}\right)}{n\sum_{j=1}^{n}\hat{v}_{1j}^2 - \left(\sum_{j=1}^{n}\hat{v}_{1j}\right)^2}$$

$$\hat{\beta}_0 = \overline{\hat{u}}_1 - \hat{\beta}_1\overline{\hat{v}}_1, \text{ 而 } \overline{\hat{u}}_1 = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_{1j}, \overline{\hat{v}}_1 = \frac{1}{n}\sum_{j=1}^{n}\hat{v}_{1j}.$$

(3.14)是大家熟知的一元线性回归方程。

在地质找矿的研究工作中，经常使用一种趋势面分析方法，这种方法是建立某地质元素 z 与地理坐标 (x,y) 之间的非线性回归方程：

$$\hat{z} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 y + \hat{\beta}_3 x^2 + \hat{\beta}_4 y^2 + \hat{\beta}_5 xy + \cdots$$

这种回归方程称为趋势面方程，利用它可以来预测元素 z 的趋势，以及利用回归残差 $e_i = z_i - \hat{z}_i$ 的大小来圈定元素 z 的异常点。然而在地质中常常要考虑一组元素的综合趋势，此时我们可以将这一组元素作为一组变量 $x^{(1)}$，把地理坐标 (x,y) 的各次幂 x, $y, x^2, y^2, xy, \cdots$. 作为另一组变量 $x^{(2)}$. 然后建立这两组变量的典型变量。譬如第一对典型变量为

$$\hat{u}_1 = \hat{l}'_1 x^{(1)}, \quad \hat{v}_1 = \hat{m}'_1 x^{(2)}$$

$\hat{u}_1$ 是一组地质元素的综合变量，$\hat{v}_1$ 是地理坐标的多项式，此时建立 $\hat{u}_1$ 与 $\hat{v}_1$ 的回归便是典型趋势面回归。它的实际应用，有兴趣的读者可参看[34]第十二章。

习 题 十 一

11.1 设两组变量 $x^{(1)} = (x_1, x_2)'$, $x^{(2)} = (x_3, x_4)'$ 有相关矩阵如下：

$$R_{11} = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}, \quad R_{22} = \begin{bmatrix} 1 & \gamma \\ \gamma & 1 \end{bmatrix}, \quad R_{12} = R'_{21} = \beta 11' = \begin{bmatrix} \beta & \beta \\ \beta & \beta \end{bmatrix}$$

且 $|\alpha| < 1$, $|\beta| < 1$, $|\gamma| < 1$, 试求它们的典型相关变量和典型相关系数。

11.2 C. R. Rao (1952) 对 25 个家庭的成年长子的头长 (x_1)、头宽 (x_2) 与成年次子的头长 (x_3)、头宽 (x_4) 进行调查，并计算了它们的协方差矩阵如下：

$$\hat{\Sigma} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{array}\begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \begin{bmatrix} 95.2933 & 52.8683 & 69.6617 & 46.1117 \\ 52.8683 & 54.3600 & 51.3117 & 35.0533 \\ 69.6617 & 51.3117 & 60.8067 & 56.5400 \\ 46.1117 & 35.0533 & 56.5400 & 45.0233 \end{bmatrix} \end{array}$$

试对 $x^{(1)} = (x_1, x_2)'$ 和 $x^{(2)} = (x_3, x_4)'$ 进行典型相关分析。

11.3 设 x 和 y 是 p_1 维和 p_2 维随机向量，且有二阶矩存在，它们的第 i 对典型变量为 $l_i' x$, $m_i' y$, ($i = 1, 2, \cdots, p_1, p_1 \leqslant p_2$) 典型相关系数为 λ_i ($i = 1, 2, \cdots, p_1$)。令 $x^* = Cx + a, y^* = Dy + b$，其中 C 和 D 分别为 $p_1 \times p_1$ 阶和 $p_2 \times p_2$ 阶非奇异阵，a 和 b 分别为 p_1 维和 p_2 维向量。试证明

(i) x^* 和 y^* 的第 $i(i = 1, 2, \cdots, p_1)$ 对典型变量为

$$C^{-1} l_i' x^*,\ D^{-1} m_i' y^*$$

(ii) $C^{-1} l_i' x^*$ 与 $D^{-1} m_i' y^*$ 的典型相关系数为 λ_i。

11.4 试证明第三节中检验假设(3.1)与检验假设(3.2)的等价性。

11.5 证明(3.5)式成立。

参 考 文 献

[1] 复旦大学编:《概率论》(第一、二册),高等教育出版社,北京,1979.

[2] 周概容:《概率论与数理统计》,高等教育出版社,北京,1984.

[3] 陈希孺:《数理统计引论》,科学出版社,北京,1981.

[4] 倪国熙: 《常用的矩阵理论和方法》,上海科技出版社,上海,1984.

[5] H.克拉美:《统计学数学方法》(魏宗舒等译),上海科技出版社,上海,1945.

[6] C.R.劳:《线性统计推断及其应用》(张燮等译),科学出版社,北京,1975.

[7] Anderson, T.W.: *An Introduction to Multivariate Statistical Analysis* (2nd Edition), Wiley, 1984.

[8] Giri, N.C.: *Multivariate Statistical Inference*, Academic press, New york, 1977.

[9] 张尧庭、方开泰:《多元统计分析引论》,科学出版社,北京,1982.

[10] M.肯德尔:《多元分析》(中科院计算中心译),科学出版社,北京,1975.

[11] 周光亚等:《多元统计分析》,地质出版社,北京,1982.

[12] 王学仁: 《地质数据的多变量统计分析》,科学出版社,北京,1982.

[13] Muirhead, R.J.: *Aspects of Multivariate Statistical Theory*, Wiley, 1982.

[14] Morrison, D. F.: *Multivariate Statistical Methods*, McGraw-Hill, Inc, 1967.

[15] Srivastava, M.S. and Khatri, C.G.: *An Introduction to Multivariate Statistical* alalysis, North-Holland, 1979.

[16] Srivastava, M.S. and Carter, E.M.: *An Introduction to Applied Multivariate Statistics*, North-Holland, 1983.

[17] 阳含熙、卢泽愚:《植物生态学的数量分类方法》,科学出版社,北京,1981.

[18] 唐守正：《多元统计分析方法》，中国林业出版社，北京，1986.

[19] 中国科学院地质研究所：《数学地质引论》，地质出版社，北京，1977.

[20] 陈希孺、王松桂：《近代回归分析》，安徽教育出版社，合肥，1987.

[21] Jöreskog, K.G. and Klovan, J.E.: *Geological Factor Analysis*. New York, 1976.

[22] 中国科学院应用数学所：《多元分析资料汇编》III、VII、VIII.

[23] Wakimoto, K.: *Tree Graph Method for Visual Representation of Multi-Dimensional Data*, Japan Statist. Soc, 27—34, 1977.

[24] Taguri, M. and Hiramatsu, M.: *Graphical Representation of Correlation Analysis of Ordered Data by Linked Vector Pattern*, Japan Statis. Soc., 17～25, 1976.

[25] Andrews, D.F., *Plots of High-Dimensional Data*, Biometrica 28, 125—36, 1972.

[26] Johonson, R.A. and Wichern, D.W. *Applied Multivariate Statistical Analysis*, 1982.

[27] Fisher, L. and Vanness, J. W., *Admissile clustering procedures*, Biometrika, 58, pp. 55—65, 1971.

[28] 王朝瑞：《图论》人民教育出版社，北京，1981.

[29] Everitt, B., *Graphical Techniques for Multivariate Data*, Heinemann, London, 1978.

[30] Seber, G.A.F., *Multivariate Observations*, Wiley, 1984.

[31] 王松桂："主成分的最优性及广义主成分估计类"，载《应用概率统计》p23—30，1985.

[32] 王松桂、林春土："主成分最优性"，载《科学通报》8(1984).

[33] 福永圭之介：《统计图形认别导论》(陶笃纯译)，科学出版社，北京，1972.

[34] 于崇文等：《数学地质方法与应用》，冶金工业出版社，北京，1980.

[35] 中山大学：《测度与概率基础》，广东科技出版社，广州，1984.

附表Ⅰ 标准正态分布 $N(0.1)$ 的分布函数值表

表中列出了 $F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$ 的值

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6430	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9993	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

x	1.282	1.645	1.960	2.326	2.576	3.090	3.291	3.991	4.417
$F(x)$	0.90	0.95	0.975	0.99	0.995	0.999	0.9995	0.99995	0.999995
$2[1-F(x)]$	0.20	0.10	0.05	0.02	0.01	0.002	0.001	0.0001	0.00001

附表Ⅱ　χ^2 分布的 $\chi^2_\alpha(n)$ 值表

表中列出了 $\chi^2(n)$ 的 α 级分位数 $\chi^2_\alpha(n)$，它满足 $P(\chi^2(n)<\chi^2_\alpha(n))=\alpha$

n \ α	0.005	0.010	0.025	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.975	0.990	0.995
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.4	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.8	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	35.7

附表III t 分布的 $t_\alpha(n)$ 值表

表中列出了 $t(n)$ 的 α 级分位数 $t_\alpha(n)$ 它满足 $P(t(n)<t_\alpha(n))=\alpha$

n \ α	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	1.638	2.353	3.182	4.541	5.841	12.941
4	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	6.859
6	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	5.405
8	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.291

附表 IV　F 分布的 $F_{\alpha}(m,n)$ 值表

表出列出了 $F(m,n)$ 的 α 级分位数 $F_{\alpha}(m,n)$，它满足 $P(F(m,n)<F_{\alpha}(m,n))=\alpha$

(2) $\alpha=0.90$。

n \ m	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	∞
1	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	61.2	61.7	62.3	62.7	63.0	63.2	63.3	63.3
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.42	9.44	9.46	9.47	9.48	9.49	9.49	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.20	5.18	5.17	5.15	5.14	5.14	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.87	3.84	3.82	3.80	3.78	3.77	3.76	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.24	3.21	3.17	3.15	3.13	3.12	3.11	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.87	2.84	2.80	2.77	2.75	2.73	2.73	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.63	2.59	2.56	2.52	2.50	2.48	2.48	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.46	2.42	2.38	2.35	2.32	2.31	2.30	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.34	2.30	2.25	2.22	2.19	2.17	2.17	2.16
10	3.28	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.24	2.20	2.16	2.12	2.09	2.07	2.06	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.17	2.12	2.08	2.04	2.00	1.99	1.98	1.97
12	31.8	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.10	2.06	2.01	1.97	1.94	1.92	1.91	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.05	2.01	1.96	1.92	1.88	1.86	1.85	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.01	1.96	1.91	1.87	1.83	1.82	1.80	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	1.97	1.92	1.87	1.83	1.79	1.77	1.76	1.76

（续表）

n \ m	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	∞
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.94	1.89	1.84	1.79	1.76	1.74	1.73	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.91	1.86	1.81	1.76	1.73	1.71	1.69	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.89	1.84	1.78	1.74	1.70	1.68	1.67	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.86	1.81	1.76	1.71	1.67	1.65	1.64	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.84	1.79	1.74	1.69	1.65	1.63	1.62	1.61
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.81	1.76	1.70	1.65	1.61	1.59	1.58	1.57
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.78	1.73	1.67	1.62	1.58	1.56	1.54	1.53
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.76	1.71	1.65	1.59	1.55	1.53	1.51	1.50
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.74	1.69	1.63	1.57	1.53	1.50	1.49	1.48
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.72	1.67	1.61	1.55	1.51	1.48	1.47	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.66	1.61	1.54	1.48	1.43	1.41	1.39	1.38
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73	1.63	1.57	1.50	1.44	1.39	1.36	1.34	1.33
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.60	1.54	1.48	1.41	1.36	1.33	1.31	1.29
80	2.77	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.71	1.68	1.57	1.51	1.44	1.38	1.32	1.28	1.26	1.24
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.70	1.66	1.56	1.49	1.42	1.35	1.29	1.26	1.23	1.21
200	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.52	1.46	1.38	1.31	1.24	1.20	1.17	1.14
500	2.72	2.31	2.10	1.96	1.86	1.79	1.73	1.68	1.64	1.61	1.50	1.44	1.36	1.28	1.21	1.16	1.12	1.09
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.49	1.42	1.34	1.26	1.18	1.13	1.08	1.00

(3) $\alpha = 0.95$,

(续表)

n \ m	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	161	200	216	225	230	234	237	239	241	242	244	245	246	247	248
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	8.67	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	5.82	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	4.58	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	3.90	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.51	2.48	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.33	2.30	2.28
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.33	2.29	2.26	2.23
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.25	2.22	2.19
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.26	2.21	2.18	2.16
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.22	2.18	2.15	2.12
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.20	2.16	2.12	2.10
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.17	2.13	2.10	2.07
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.15	2.11	2.07	2.05
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.13	2.09	2.05	2.03
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.11	2.07	2.04	2.01

(续表)

n \ m	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.09	2.05	2.02	1.99
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.08	2.04	2.00	1.97
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.06	2.02	1.99	1.96
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.05	2.01	1.97	1.94
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.04	1.99	1.96	1.93
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.07	2.01	1.97	1.94	1.91
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	2.05	1.99	1.95	1.92	1.89
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	2.03	1.98	1.93	1.90	1.87
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.02	1.96	1.92	1.88	1.85
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.95	1.90	1.87	1.84
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	1.99	1.93	1.89	1.86	1.83
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	1.89	1.92	1.88	1.84	1.81
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04	1.97	1.91	1.87	1.83	1.80
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03	1.96	1.90	1.86	1.82	1.79
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.89	1.85	1.81	1.78
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.86	1.82	1.78	1.75
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.88	1.82	1.77	1.73	1.70
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.79	1.75	1.71	1.68
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.96	1.91	1.83	1.77	1.72	1.69	1.65
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.64	1.89	1.82	1.76	1.71	1.67	1.64
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.80	1.74	1.69	1.66	1.62
300	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.78	1.72	1.68	1.64	1.61
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.77	1.71	1.66	1.62	1.59
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.70	1.65	1.61	1.58
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.69	1.64	1.60	1.57

(3) $\alpha=0.95$ (续)

(续表)

n \ m	22	24	26	28	30	35	40	45	50	60	80	100	200	500	∞
1	249	249	249	250	250	251	251	251	252	252	252	253	254	254	254
2	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5
3	8.65	8.64	8.63	8.62	8.62	8.60	8.59	8.59	8.58	8.57	8.56	8.55	8.54	8.53	8.53
4	5.79	5.77	5.76	5.75	5.75	5.73	5.72	5.71	5.70	5.69	5.67	5.66	5.65	5.64	5.63
5	4.54	4.53	4.52	4.50	4.50	4.48	4.46	4.45	4.44	4.43	4.41	4.41	4.39	4.37	4.37
6	3.86	3.84	3.83	3.82	3.81	3.79	3.77	3.76	3.75	3.74	3.72	3.71	3.69	3.68	3.67
7	3.43	3.41	3.40	3.39	3.38	3.36	3.34	3.33	3.32	3.30	3.29	3.27	3.25	3.24	3.23
8	3.13	3.12	3.10	3.09	3.08	3.06	3.04	3.03	3.02	3.01	2.99	2.97	2.95	2.94	2.93
9	2.92	2.90	2.89	2.87	2.86	2.84	2.83	2.81	2.80	2.79	2.77	2.76	2.73	2.72	2.71
10	2.75	2.74	2.72	2.71	2.70	2.68	2.66	2.65	2.64	2.62	2.60	2.59	2.56	2.55	2.54
11	2.63	2.61	2.59	2.58	2.57	2.55	2.53	2.52	2.51	2.49	2.47	2.46	2.43	2.42	2.40
12	2.52	2.51	2.49	2.48	2.47	2.44	2.43	2.41	2.40	2.38	2.36	2.35	2.32	2.31	2.30
13	2.44	2.42	2.41	2.39	2.38	2.36	2.34	2.33	2.31	2.30	2.27	2.26	2.23	2.22	2.21
14	2.37	2.35	2.33	2.32	2.31	2.28	2.27	2.25	2.24	2.22	2.20	2.19	2.16	2.14	2.13
15	2.31	2.29	2.27	2.26	2.25	2.22	2.20	2.19	2.18	2.16	2.14	2.12	2.10	2.08	2.07
16	2.25	2.24	2.22	2.21	2.19	2.17	2.15	2.14	2.12	2.11	2.08	2.07	2.04	2.02	2.01
17	2.21	2.19	2.17	2.16	2.15	2.12	2.10	2.09	2.08	2.06	2.03	2.02	1.99	1.97	1.96
18	2.17	2.15	2.13	2.12	2.11	2.08	2.06	2.05	2.04	2.02	1.99	1.98	1.95	1.93	1.92
19	2.13	2.11	2.10	2.08	2.07	2.05	2.03	2.01	2.00	1.98	1.96	1.94	1.91	1.89	1.88
20	2.10	2.08	2.07	2.05	2.04	2.01	1.99	1.98	1.97	1.95	1.92	1.91	1.88	1.86	1.84
21	2.07	2.05	2.04	2.02	2.01	1.98	1.96	1.95	1.94	1.92	1.89	1.88	1.84	1.82	1.81
22	2.05	2.03	2.01	2.00	1.98	1.96	1.94	1.92	1.91	1.89	1.86	1.85	1.82	1.80	1.78
23	2.02	2.00	1.99	1.97	1.96	1.93	1.91	1.90	1.88	1.86	1.84	1.82	1.79	1.77	1.76
24	2.00	1.98	1.97	1.95	1.94	1.91	1.89	1.88	1.86	1.84	1.82	1.80	1.77	1.75	1.73
25	1.98	1.96	1.95	1.93	1.92	1.89	1.87	1.86	1.84	1.82	1.80	1.78	1.75	1.73	1.71

（续表）

n \ m	22	24	26	28	30	35	40	45	50	60	80	100	200	500	∞
26	1.97	1.95	1.93	1.91	1.90	1.87	1.85	1.84	1.82	1.80	1.78	1.76	1.73	1.71	1.69
27	1.95	1.93	1.91	1.90	1.88	1.86	1.84	1.82	1.81	1.79	1.76	1.74	1.71	1.69	1.67
28	1.93	1.91	1.90	1.88	1.87	1.84	1.82	1.80	1.79	1.77	1.74	1.73	1.69	1.67	1.65
29	1.92	1.90	1.88	1.87	1.85	1.83	1.81	1.79	1.77	1.75	1.73	1.71	1.67	1.65	1.64
30	1.91	1.89	1.87	1.85	1.84	1.81	1.79	1.77	1.76	1.74	1.71	1.70	1.66	1.64	1.62
32	1.88	1.86	1.85	1.83	1.82	1.79	1.77	1.75	1.74	1.71	1.69	1.67	1.63	1.61	1.59
34	1.86	1.84	1.82	1.80	1.80	1.77	1.75	1.73	1.71	1.69	1.66	1.65	1.61	1.59	1.57
36	1.85	1.82	1.81	1.79	1.78	1.75	1.73	1.71	1.69	1.67	1.64	1.62	1.59	1.56	1.55
38	1.83	1.81	1.79	1.77	1.76	1.73	1.71	1.69	1.68	1.65	1.62	1.61	1.57	1.54	1.53
40	1.81	1.79	1.77	1.76	1.74	1.72	1.69	1.67	1.66	1.64	1.61	1.59	1.55	1.53	1.51
42	1.80	1.78	1.76	1.74	1.73	1.70	1.68	1.66	1.65	1.62	1.59	1.57	1.53	1.51	1.49
44	1.79	1.77	1.75	1.73	1.72	1.69	1.67	1.65	1.63	1.61	1.58	1.56	1.52	1.49	1.48
46	1.78	1.76	1.74	1.72	1.71	1.68	1.65	1.64	1.62	1.60	1.57	1.55	1.51	1.48	1.46
48	1.77	1.75	1.73	1.71	1.70	1.67	1.64	1.62	1.61	1.59	1.56	1.54	1.49	1.47	1.45
50	1.76	1.74	1.72	1.70	1.69	1.66	1.63	1.61	1.60	1.58	1.54	1.52	1.48	1.46	1.44
60	1.72	1.70	1.68	1.66	1.65	1.62	1.59	1.57	1.56	1.53	1.50	1.48	1.44	1.41	1.39
80	1.68	1.65	1.63	1.62	1.60	1.57	1.54	1.52	1.51	1.48	1.45	1.43	1.38	1.35	1.32
100	1.65	1.63	1.61	1.59	1.57	1.54	1.52	1.49	1.48	1.45	1.41	1.39	1.34	1.31	1.28
125	1.63	1.60	1.58	1.57	1.55	1.52	1.49	1.47	1.45	1.42	1.39	1.36	1.31	1.27	1.25
150	1.61	1.59	1.57	1.55	1.53	1.50	1.48	1.45	1.44	1.41	1.37	1.34	1.29	1.25	1.22
200	1.60	1.57	1.55	1.53	1.52	1.48	1.46	1.43	1.41	1.39	1.35	1.32	1.26	1.22	1.19
300	1.58	1.55	1.53	1.51	1.50	1.46	1.43	1.41	1.39	1.36	1.32	1.30	1.23	1.19	1.15
500	1.56	1.54	1.52	1.50	1.48	1.45	1.42	1.40	1.38	1.34	1.30	1.28	1.21	1.16	1.11
1000	1.55	1.53	1.51	1.49	1.47	1.44	1.41	1.38	1.36	1.33	1.29	1.26	1.19	1.13	1.08
∞	1.54	1.52	1.50	1.48	1.46	1.42	1.39	1.37	1.35	1.32	1.27	1.24	1.17	1.11	1.00

(4) $\alpha=0.99$

(续表)

n \ m	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	405	500	540	563	576	586	593	598	602	606	611	614	617	619	621
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.8	26.8	26.7
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.2	14.1	14.0
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.77	9.68	9.61	9.55
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.60	7.52	7.45	7.40
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.36	6.27	6.21	6.16
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.56	5.48	5.41	5.36
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	5.00	4.92	4.68	4.81
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.60	4.52	4.46	4.41
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.29	4.21	4.15	4.10
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.05	3.97	3.91	3.86
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.86	3.78	3.71	3.66
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.70	3.62	3.56	3.51
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.56	3.49	3.42	3.37
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.45	3.37	3.31	3.26
17	8.40	6.11	5.18	4.67	4.34	4.11	3.93	3.73	3.68	3.59	3.46	3.35	3.27	3.21	3.16
18	8.29	6.01	5.09	4.58	4.25	4.00	3.84	3.71	3.60	3.51	3.37	3.27	3.19	3.31	3.08
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.19	3.12	3.05	3.00
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.13	3.05	2.99	2.94
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.07	2.99	2.93	2.88
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	3.02	2.94	2.88	2.83
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.97	2.89	2.83	2.78
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.93	2.85	2.79	2.74
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.89	2.81	2.75	2.70

（续表）

n \ m	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.86	2.78	2.72	2.66
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.82	2.75	2.68	2.63
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.79	2.72	2.65	2.60
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.77	2.69	2.62	2.57
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.74	2.66	2.60	2.55
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93	2.80	2.70	2.62	2.55	2.50
34	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89	2.76	2.66	2.58	2.51	2.46
36	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86	2.72	2.62	2.54	2.48	2.43
38	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83	2.69	2.59	2.51	2.45	2.40
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.56	2.48	2.42	2.37
42	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86	2.78	2.64	2.54	2.46	2.40	2.34
44	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84	2.75	2.62	2.52	2.44	2.37	2.32
46	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82	2.73	2.60	2.50	2.42	2.35	2.30
48	7.20	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80	2.72	2.58	2.48	2.40	2.33	2.28
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.79	2.70	2.56	2.46	2.38	2.32	2.27
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.39	2.31	2.25	2.20
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.42	2.31	2.23	2.17	2.12
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.37	2.26	2.19	2.12	2.07
125	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.66	2.55	2.47	2.33	2.23	2.15	2.08	2.03
150	6.81	4.75	3.92	3.45	3.14	2.92	2.76	2.63	2.53	2.44	2.31	2.20	2.12	2.06	2.00
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.27	2.17	2.09	2.02	1.97
300	6.72	4.68	3.85	3.38	3.08	2.86	2.70	2.57	2.47	2.38	2.24	2.14	2.06	1.99	1.94
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.22	2.12	2.04	1.97	1.92
1000	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.20	2.10	2.02	1.95	1.90
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.08	2.00	1.93	1.88

（4）$\alpha=0.99$ （续表）

n \ m	22	24	26	28	30	35	40	45	50	60	80	100	200	500	∞
1	622	623	624	625	626	628	629	630	630	631	633	633	635	636	637
2	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
3	26.6	26.6	26.6	26.5	26.5	26.5	26.4	26.4	26.4	26.3	26.3	26.2	26.2	26.1	26.1
4	14.0	13.9	13.9	13.9	13.8	13.8	13.7	13.7	13.7	13.7	13.6	13.6	13.5	13.5	13.5
5	9.51	9.47	9.43	9.40	9.38	9.33	9.29	9.26	9.24	9.20	9.16	9.13	9.08	9.04	9.02
6	7.35	7.31	7.28	7.25	7.23	7.18	7.14	7.11	7.09	7.06	7.01	6.99	6.93	6.90	6.88
7	6.11	6.07	6.04	6.02	5.99	5.94	5.91	5.88	5.86	5.82	5.78	5.75	5.70	5.67	5.65
8	5.32	5.28	5.25	5.22	5.20	5.15	5.12	5.00	5.07	5.03	4.99	4.96	4.91	4.88	4.86
9	4.77	4.73	4.70	4.67	4.65	4.60	4.57	4.51	4.52	4.48	4.44	4.42	4.36	4.33	4.31
10	4.36	4.33	4.30	4.27	4.25	4.20	4.17	4.14	4.12	4.08	4.04	4.01	3.96	3.93	3.91
11	4.06	4.02	3.99	3.96	3.94	3.89	3.86	3.83	3.81	3.78	3.73	3.71	3.66	3.62	3.60
12	3.82	3.78	3.75	3.72	3.70	3.65	3.62	3.59	3.57	3.54	3.49	3.47	3.41	3.38	3.36
13	3.62	3.59	3.56	3.53	3.51	3.46	3.43	3.40	3.38	3.34	3.30	3.27	3.22	3.19	3.17
14	3.46	3.43	3.40	3.37	3.35	3.30	3.27	3.24	3.22	3.18	3.14	3.11	3.06	3.03	3.00
15	3.33	3.29	3.26	3.24	3.21	3.17	3.13	3.10	3.08	3.05	3.00	2.98	2.92	2.89	2.87
16	3.22	3.18	3.15	3.12	3.10	3.05	3.02	2.99	2.97	2.93	2.89	2.86	2.81	2.78	2.75
17	3.12	3.08	3.05	3.03	3.00	2.96	2.92	2.89	2.87	2.83	2.79	2.76	2.71	2.68	2.65
18	3.03	3.00	2.97	2.94	2.92	2.87	2.84	2.81	2.78	2.75	2.70	2.68	2.62	2.59	2.57
19	2.96	2.92	2.89	2.87	2.84	2.80	2.76	2.73	2.71	2.67	2.63	2.60	2.55	2.51	2.49
20	2.90	2.86	2.83	2.80	2.78	2.73	2.69	2.67	2.64	2.61	2.56	2.54	2.48	2.44	2.42
21	2.84	2.80	2.77	2.74	2.72	2.67	2.64	2.61	2.58	2.55	2.50	2.48	2.42	2.38	2.36
22	2.78	2.75	2.72	2.69	2.67	2.62	2.58	2.55	2.53	2.50	2.45	2.42	2.36	2.33	2.31
23	2.74	2.70	2.67	2.64	2.62	2.57	2.54	2.51	2.48	2.45	2.40	2.37	2.32	2.28	2.26
24	2.70	2.66	2.63	2.60	2.58	2.53	2.49	2.46	2.44	2.40	2.36	2.33	2.27	2.24	2.21
25	2.86	2.62	2.59	2.56	2.54	2.49	2.45	2.42	2.40	2.36	2.32	2.29	2.23	2.19	2.17

（续表）

n \ m	22	24	26	28	30	35	40	45	50	60	80	100	200	500	∞
26	2.62	2.58	2.55	2.53	2.50	2.45	2.42	2.39	2.36	2.33	2.28	2.25	2.19	2.16	2.13
27	2.59	2.55	2.52	2.49	2.47	2.42	2.38	2.35	2.33	2.29	2.25	2.22	2.16	2.12	2.10
28	2.56	2.52	2.49	2.46	2.44	2.39	2.35	2.32	2.30	2.26	2.22	2.19	2.13	2.09	2.06
29	2.53	2.46	2.46	2.44	2.41	2.36	2.33	2.30	2.27	2.23	2.19	2.16	2.10	2.06	2.03
30	2.51	2.47	2.44	2.41	2.39	2.34	2.30	2.27	2.25	2.21	2.16	2.13	2.07	2.03	2.01
32	2.46	2.42	2.39	2.36	2.34	2.29	2.25	2.22	2.20	2.16	2.11	2.08	2.02	1.98	1.96
34	2.42	2.38	2.35	2.32	2.30	2.25	2.21	2.18	2.16	2.12	2.07	2.04	1.98	1.94	1.91
36	2.38	2.35	2.32	2.29	2.26	2.21	2.17	2.14	2.12	2.08	2.03	2.00	1.94	1.90	1.87
38	2.35	2.32	2.28	2.26	2.23	2.18	2.14	2.11	2.09	2.05	2.00	1.97	1.90	1.86	1.84
40	2.33	2.29	2.26	2.23	2.20	2.15	2.11	2.08	2.06	2.02	1.97	1.94	1.87	1.83	1.80
42	2.30	2.26	2.23	2.20	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.91	1.85	1.80	1.78
44	2.28	2.24	2.21	2.18	2.15	2.10	2.06	2.03	2.01	1.97	1.92	1.89	1.82	1.78	1.75
46	2.26	2.22	2.19	2.16	2.13	2.08	2.04	2.01	1.99	1.95	1.90	1.86	1.80	1.75	1.73
48	2.24	2.20	2.17	2.14	2.12	2.06	2.02	1.99	1.97	1.93	1.88	1.84	1.78	1.73	1.70
50	2.22	2.18	2.15	2.12	2.10	2.05	2.01	1.97	1.95	1.91	1.86	1.82	1.76	1.71	1.68
60	2.15	2.12	2.08	2.05	2.03	1.98	1.94	1.90	1.88	1.84	1.78	1.75	1.68	1.63	1.60
80	2.07	2.03	2.00	1.97	1.94	1.89	1.85	1.81	1.79	1.75	1.69	1.66	1.58	1.53	1.49
100	2.02	1.98	1.94	1.92	1.89	1.84	1.80	1.76	1.73	1.69	1.63	1.60	1.52	1.47	1.43
125	1.98	1.94	1.91	1.88	1.85	1.80	1.76	1.72	1.69	1.65	1.59	1.55	1.47	1.41	1.37
150	1.96	1.92	1.88	1.85	1.83	1.77	1.73	1.69	1.66	1.62	1.56	1.52	1.43	1.38	1.33
200	1.93	1.89	1.85	1.82	1.79	1.74	1.69	1.66	1.63	1.58	1.52	1.48	1.39	1.33	1.28
300	1.89	1.85	1.82	1.79	1.76	1.71	1.66	1.62	1.59	1.55	1.48	1.44	1.35	1.28	1.22
500	1.87	1.83	1.79	1.76	1.74	1.68	1.63	1.60	1.56	1.52	1.45	1.41	1.31	1.23	1.16
1000	1.85	1.81	1.77	1.74	1.72	1.66	1.61	1.57	1.54	1.50	1.43	1.38	1.28	1.19	1.11
∞	1.83	1.79	1.76	1.72	1.70	1.64	1.59	1.55	1.52	1.47	1.40	1.36	1.25	1.15	1.00

附表V θ_{max} 分布的 $\theta_{max}(p,n,m,\alpha)$ 值表*

表中列出了 $\theta_{max}(p,n,m)$ 的 α 级分位数 $\theta_{max}(p,n,m,\alpha)$

它满足 $P(\theta_{max}(p,n,m)<\theta_{max}(p,n,m,\alpha))=\alpha$

$p=2,\alpha=0.90$			m			
n	1	2	3	4	5	6
3	0.8000	0.9487	0.9655	0.9740	0.9791	0.9825
5	0.6838	0.7950	0.8463	0.8767	0.8908	0.9112
7	0.5358	0.6628	0.7307	0.7746	0.8058	0.8292
9	0.4377	0.5632	0.6366	0.6805	0.7244	0.7535
11	0.3690	0.4880	0.5618	0.6146	0.6551	0.6873
13	0.3137	0.4298	0.5017	0.5548	0.5965	0.6304
15	0.2803	0.3837	0.4527	0.5049	0.5467	0.5813
17	0.2501	0.3464	0.4122	0.4630	0.5043	0.5389
19	0.2257	0.3155	0.3782	0.4273	0.4676	0.5019
21	0.2051	0.2897	0.3493	0.3465	0.4359	0.4695
25	0.1746	0.2488	0.3028	0.3465	0.3834	0.4155
31	0.1423	0.2053	0.2523	0.2911	0.3245	0.3541
35	0.1261	0.1838	0.2270	0.2630	0.2943	0.3222
41	0.1087	0.1589	0.1972	0.2296	0.2581	0.2837
61	0.0739	0.1093	0.1372	0.1613	0.1830	0.2028
101	0.0450	0.0673	0.0853	0.1011	0.1155	0.1290

$p=2,\alpha=0.90$				m			
n	7	8	9	10	12	15	20
3	0.9851	0.9869	0.9884	0.9895	0.9913	0.9930	0.9947
5	0.9221	0.9306	0.9374	0.9430	0.9516	0.9605	0.9698
7	0.8474	0.8620	0.8741	0.8842	0.9002	0.9173	0.9356
9	0.7768	0.7959	0.8120	0.8257	0.8477	0.8703	0.8987
11	0.7138	0.7360	0.7549	0.7711	0.7978	0.8278	0.8618
13	0.6587	0.6827	0.7035	0.7216	0.7517	0.7862	0.8261
15	0.6106	0.6357	0.6577	0.0770	0.7096	0.7474	0.7922
17	0.5685	0.5942	0.6169	0.6370	0.6712	0.7116	0.7603
19	0.5315	0.5575	0.5805	0.6011	0.6364	0.6787	0.7303
21	0.4989	0.5248	0.5479	0.5687	0.6047	0.6482	0.7023
25	0.4439	0.4692	0.4921	0.5129	0.5494	0.5943	0.6515
31	0.3805	0.4045	0.4264	0.4405	0.4824	0.5276	0.5869
35	0.3473	0.3703	0.3914	0.4109	0.4459	0.4906	0.5501
41	0.3070	0.3284	0.3483	0.3668	0.4003	0.4436	0.5025
61	0.2211	0.2383	0.2544	0.2097	0.2979	0.3357	0.3891
101	0.1416	0.1536	0.1651	0.1761	0.1968	0.2253	0.2674

(续表)

$p=3, \alpha=0.90$			m			
n	1	2	3	4	5	6
4	0.9322	0.9655	0.9769	0.9826	0.9801	0.9884
6	0.7586	0.8463	0.8857	0.9081	0.9238	0.9346
8	0.6218	0.7307	0.7871	0.8229	0.8480	0.8607
10	0.5230	0.6366	0.7008	0.7440	0.7757	0.8000
12	0.4500	0.5618	0.6287	0.6757	0.7112	0.7392
14	0.3944	0.5017	0.5681	0.0172	0.6548	0.6851
16	0.3509	0.4527	0.5186	0.5672	0.6057	0.6372
18	0.3158	0.4122	0.4760	0.5242	0.5629	0.5950
20	0.2871	0.3782	0.4397	0.4870	0.3254	0.5576
26	0.2254	0.3028	0.3574	0.4007	0.4370	0.4681
30	0.1971	0.2671	0.3175	0.3581	0.3925	0.4224
36	0.1658	0.2270	0.2718	0.3087	0.3682	0.3682
40	0.1499	0.2062	0.2480	0.2526	0.3125	0.3390
60	0.1013	0.1416	0.1723	0.1984	0.2216	0.2426
100	0.0615	0.0869	0.1069	0.1242	0.1399	0.1543

$p=3, \alpha=0.90$			m				
n	7	8	9	10	12	15	20
4	0.9900	0.9913	0.9922	0.9930	0.9942	0.9953	0.9964
6	0.9427	0.9490	0.9541	0.9582	0.9645	0.9711	0.9779
8	0.8812	0.8929	0.9024	0.9103	0.9228	0.9362	0.9504
10	0.8195	0.8353	0.8486	0.8598	0.8779	0.8976	0.9192
12	0.7620	0.7810	0.7971	0.8109	0.8335	0.8586	0.8869
14	0.7101	0.7313	0.7494	0.7652	0.7912	0.8208	0.8549
16	0.6637	0.6863	0.7058	0.7230	0.7518	0.7849	0.8238
18	0.6222	0.6457	0.6663	0.6845	0.7152	0.7512	0.7948
20	0.5852	0.6093	0.6304	0.6493	0.6814	0.7196	0.7658
26	0.4954	0.5196	0.5413	0.5610	0.5952	0.6370	0.6897
30	0.4489	0.4726	0.4941	0.5136	0.5481	0.5909	0.6459
36	0.3932	0.4158	0.4364	0.4555	0.4894	0.5323	0.5889
40	0.3630	0.3848	0.4048	0.4233	0.4566	0.4991	0.5558
60	0.2619	0.2798	0.2965	0.3122	0.3411	0.3793	0.4328
100	0.1679	0.1806	0.1927	0.2043	0.2259	0.2554	0.2981

(续表)

$p=4,\alpha=0.90$	m					
n	1	2	3	4	5	6
5	0.9487	0.9740	0.9826	0.9869	0.9895	0.9913
7	0.8042	0.8767	0.9087	0.9272	0.9394	0.9480
9	0.6795	0.7746	0.8229	0.8532	0.8744	0.8900
11	0.5839	0.6805	0.7440	0.7820	0.8095	0.8306
13	0.5103	0.6146	0.6757	0.7180	0.7498	0.7746
15	0.4526	0.5548	0.6172	0.6619	0.6961	0.7235
17	0.4002	0.5049	0.5672	0.6128	0.6485	0.6774
19	0.3654	0.4630	0.5242	0.5700	0.6063	0.6361
21	0.3368	0.4272	0.4870	0.5323	0.5688	0.5991
25	0.2875	0.3698	0.4260	0.4096	0.5054	0.5358
31	0.2356	0.3075	0.3581	0.3985	0.4323	0.4615
35	0.2102	0.3763	0.3236	0.3617	0.3940	0.4222
41	0.1810	0.2398	0.2826	0.3176	0.3476	0.3741
61	0.1236	0.1663	0.1984	0.2254	0.2492	0.2706
101	0.0756	0.1030	0.1242	0.1425	0.1588	0.1738

$p=4,\alpha=0.90$	m						
n	7	8	9	10	12	15	20
5	0.9925	0.9934	0.9942	0.9947	0.9956	0.9965	0.9974
7	0.9545	0.9595	0.9636	0.9668	0.9719	0.9771	0.9826
9	0.9022	0.9118	0.9197	0.9263	0.9367	0.9777	0.9594
11	0.8473	0.8610	0.8723	0.8819	0.8973	0.9140	0.9323
13	0.7947	0.8114	0.8255	0.8376	0.8572	0.8791	0.9035
15	0.7460	0.7650	0.7812	0.7952	0.8183	0.8445	0.8744
17	0.7016	0.7221	0.7400	0.7554	0.7813	0.8110	0.8457
19	0.6614	0.6830	0.7020	0.7185	0.7466	0.7792	0.8178
21	0.6250	0.6474	0.6670	0.6845	0.7141	0.7490	0.7910
25	0.5621	0.5851	0.6056	0.6240	0.6557	0.6939	0.7410
31	0.4873	0.5102	0.5309	0.5496	0.5826	0.6232	0.6749
35	0.4472	0.4697	0.4901	0.4901	0.5087	0.5417	0.6363
41	0.3978	0.4198	0.4390	0.4572	0.4896	0.5308	0.5854
61	0.2902	0.3082	0.3251	0.3408	0.3697	0.4076	0.4602
101	0.1878	0.2009	0.2134	0.2252	0.2472	0.2772	0.3208

（续表）

$p=5, \alpha=0.90$	m				
n	1	2	3	4	5
6	0.9587	0.9791	0.9861	0.9895	0.9916
8	0.8351	0.8968	0.9238	0.9394	0.9496
10	0.7214	0.8058	0.8480	0.8744	0.8927
12	0.6302	0.7244	0.7757	0.8095	0.8339
14	0.5577	0.6551	0.7112	0.7498	0.7784
16	0.4994	0.5965	0.6548	0.6961	0.7276
18	0.4517	0.5467	0.6057	0.6485	0.6817
20	0.4122	0.5043	0.5629	0.6063	0.6404
26	0.3260	0.4080	0.4631	0.5054	0.5400
30	0.2860	0.3616	0.4136	0.4543	0.4881
36	0.2413	0.3087	0.3561	0.3940	0.4259
40	0.2186	0.2811	0.3258	0.3618	0.3924
60	0.1485	0.1943	0.2282	0.2565	0.2811
100	0.0804	0.1200	0.1425	0.1618	0.1789

$p=2, \alpha=0.95$	m					
n	1	2	3	4	5	6
3	0.9500	0.9748	0.9830	0.9873	0.9898	0.9915
5	0.7764	0.8577	0.8943	0.9155	0.9296	0.9396
7	0.6316	0.7370	0.7919	0.8269	0.8514	0.8697
9	0.5271	0.0383	0.7017	0.7446	0.7761	0.8003
11	0.4507	0.5603	0.6267	0.6735	0.7090	0.7371
13	0.3930	0.4981	0.5646	0.6131	0.6507	0.6811
15	0.3438	0.4479	0.5130	0.5617	0.6002	0.6319
17	0.3123	0.4065	0.4696	0.5177	0.5564	0.5886
19	0.2831	0.3720	0.4328	0.4798	0.5182	0.5505
21	0.2589	0.3427	0.4011	0.4469	0.5167	0.5167
25	0.2209	0.2961	0.3497	0.3927	0.4287	0.4598
31	0.1810	0.2457	0.2931	0.3319	0.3650	0.3441
35	0.1616	0.2206	0.2645	0.3007	0.3320	0.3597
41	0.1391	0.1913	0.2306	0.2635	0.2922	0.3178
61	0.0950	0.1324	0.1615	0.1864	0.2085	0.2287
101	0.0582	0.0820	0.1009	0.1174	0.1325	0.1464

（续表）

$p=2,\alpha=0.95$				m			
n	7	8	9	10	12	15	20
3	0.9927	0.9936	0.9943	0.9949	0.9957	0.9966	0.9974
5	0.9471	0.9529	0.9576	0.9613	0.9672	0.9733	0.9796
7	0.8839	0.8952	0.9045	0.9123	0.9246	0.9376	0.9515
9	0.8197	0.8355	0.8487	0.8599	0.8780	0.8976	0.9192
11	0.7600	0.7791	0.7953	0.8091	0.8319	0.8572	0.8858
13	0.7063	0.7276	0.7459	0.7618	0.7882	0.8181	0.8526
15	0.6585	0.6813	0.7010	0.7184	0.7474	0.7811	0.8206
17	0.6160	0.6397	0.6604	0.678	0.7099	0.7403	0.7899
19	0.5782	0.6024	0.6238	0.6428	0.6753	0.7139	0.7609
21	0.5445	0.5689	0.5906	0.6100	0.6435	0.6838	0.7334
25	0.4872	0.5114	0.5333	0.5530	0.5875	0.6298	0.6832
31	0.4200	0.4434	0.4647	0.4842	0.5187	0.5620	0.6183
35	0.3845	0.4071	0.4277	0.4468	0.4808	0.5240	0.5810
41	0.3411	0.3623	0.3820	0.4002	0.4331	0.4754	0.5324
61	0.2474	0.2648	0.2810	0.2964	0.3247	0.3623	0.4152
101	0.1594	0.1718	0.1835	0.1947	0.2158	0.2447	0.2872

$p=3,\alpha=0.95$				m		
n	1	2	3	4	5	6
4	0.9664	0.9830	0.9887	0.9915	0.9932	0.9943
6	0.8318	0.8943	0.9218	0.9378	0.9482	0.9556
8	0.7040	0.7919	0.8365	0.8646	0.8842	0.8986
10	0.6039	0.7017	0.7560	0.7922	0.8185	0.8386
12	0.5266	0.6267	0.6857	0.7266	0.7573	0.7815
14	0.4660	0.5646	0.6254	0.6689	0.7024	0.7292
16	0.4174	0.5130	0.5739	0.6185	0.6535	0.6820
18	0.3779	0.4696	0.5296	0.5745	0.6103	0.6398
20	0.3450	0.4328	0.4914	0.5359	0.5718	0.6019
26	0.2732	0.3497	0.4030	0.4450	0.4798	0.5096
30	0.2400	0.3098	0.3595	0.3993	0.4328	0.4018
36	0.2028	0.2645	0.3093	0.3458	0.4043	0.4043
40	0.1838	0.2409	0.2828	0.3173	0.3470	0.3732
60	0.1611	0.1665	0.1979	0.2245	0.2479	0.2691
100	0.0763	0.1028	0.1235	0.1414	0.1575	0.1723

$p=3,\ \alpha=0.95$				m			
n	7	8	9	10	12	15	20
4	0.9951	0.9957	0.9962	0.9966	0.9972	0.9977	0.9983
6	0.9612	0.9655	0.9689	0.9717	0.9760	0.9805	0.9851
8	0.9098	0.9188	0.9261	0.9322	0.9417	0.9519	0.9627
10	0.8546	0.8076	0.8784	0.8876	0.9022	0.9182	0.9356
12	0.8010	0.8172	0.8309	0.8426	0.8017	0.8828	0.9065
14	0.7512	0.7698	0.7857	0.7994	0.8221	0.8477	0.8770
16	0.7058	0.7261	0.7436	0.7589	0.7844	0.8137	0.8478
18	0.6647	0.6861	0.7048	0.7212	0.7489	0.7812	0.8195
20	0.6275	0.6497	0.6692	0.6864	0.7158	0.7505	0.7923
26	0.5356	0.5586	0.5792	0.5977	0.6299	0.6690	0.7179
30	0.4873	0.5101	0.5307	0.5493	0.5822	0.6227	0.6744
36	0.4287	0.4507	0.4709	0.4893	0.5220	0.5633	0.6172
40	0.3968	0.4181	0.4377	0.4558	0.4881	0.5292	0.5838
60	0.2884	0.3064	0.3231	0.3387	0.3674	0.4052	0.4578
100	0.1861	0.1991	0.2114	0.2231	0.2450	0.2747	0.3145

$p=4,\ \alpha=0.95$				m		
n	1	2	3	4	5	6
5	0.9748	0.9873	0.9915	0.9936	0.9949	0.9957
7	0.8647	0.9155	0.9378	0.9505	0.9589	0.9648
9	0.7514	0.8268	0.8646	0.8882	0.9045	0.9166
11	0.6574	0.7446	0.7922	0.8236	0.8463	0.8636
13	0.5818	0.6735	0.7266	0.7632	0.7904	0.8116
15	0.5207	0.6131	0.6689	0.7085	0.7387	0.7628
17	0.4707	0.5617	0.6185	0.6599	0.6920	0.7180
19	0.4291	0.5177	0.5745	0.6166	0.0499	0.6771
21	0.3942	0.4798	0.5359	0.5782	0.6120	0.6400
26	0.3387	0.4181	0.4718	0.5133	0.5472	0.5758
31	0.2794	0.3500	0.3993	0.4384	0.4711	0.4992
35	0.2501	0.3155	0.3620	0.3993	0.4307	0.4581
41	0.2161	0.2748	0.3173	0.3519	0.3814	0.4074
61	0.1486	0.1920	0.2245	0.2517	0.2755	0.2970
101	0.0914	0.1197	0.1414	0.1600	0.1767	0.1920

$p=4,\ \alpha=0.95$				m			
n	7	8	9	10	12	15	20
5	0.9963	0.9968	0.9972	0.9974	0.9979	0.9983	0.9987
7	0.9692	0.9726	0.9754	0.9776	0.9810	0.9846	0.9882
9	0.9259	0.9333	0.9393	0.9443	0.9522	0.9606	0.9695
11	0.8773	0.8884	0.8976	0.9054	0.9178	0.9313	0.9460
13	0.8287	0.8429	0.8548	0.8470	0.8815	0.8998	0.9202
15	0.7825	0.7996	0.8131	0.8253	0.8453	0.8678	0.8985
17	0.7346	0.7579	0.7737	0.7874	0.8103	0.8364	0.8667
19	0.7000	0.7197	0.7367	0.7517	0.7768	0.8060	0.8404
21	0.6038	0.6844	0.7024	0.7183	0.7452	0.7760	0.8147
26	0.6004	0.0220	0.6412	0.6583	0.6877	0.7229	0.7663
31	0.5238	0.5457	0.5654	0.5832	0.6144	0.6527	0.7012
35	0.4822	0.5039	0.5235	0.5414	0.5729	0.6121	0.6628
41	0.4306	0.4516	0.4701	0.4883	0.5197	0.5593	0.6116
61	0.3164	0.3344	0.3511	0.3668	0.3952	0.4326	0.4842
101	0.2062	0.2195	0.2320	0.2439	0.2662	0.2962	0.3897

$p=5,\ \alpha=0.95$			m		
n	1	2	3	4	5
6	0.9797	0.9898	0.9932	0.9949	0.9960
8	0.8866	0.9296	0.9482	0.9589	0.9659
10	0.7852	0.8514	0.8842	0.9045	0.9186
12	0.0974	0.7761	0.8185	0.8463	0.8663
14	0.6245	0.7090	0.7574	0.7904	0.8148
16	0.5641	0.6507	0.7024	0.7387	0.7663
18	0.5137	0.6002	0.6535	0.6920	0.7217
20	0.4713	0.5564	0.6103	0.6499	0.6810
26	0.3769	0.4550	0.5072	0.6472	0.5796
30	0.3322	0.4052	0.4552	0.4941	0.5262
36	0.2819	0.3477	0.3940	0.4307	0.4616
40	0.2560	0.3176	0.3614	0.3966	0.4264
60	0.1752	0.2212	0.2552	0.2834	0.3079
100	0.1074	0.1375	0.1604	0.1800	0.1973

$p=2,\ \alpha=0.99$			m			
n	1	2	3	4	5	6
3	0.9900	0.9950	0.9967	0.9975	0.9980	0.9983
5	0.9000	0.9378	0.9542	0.9636	0.9698	0.9742
7	0.7846	0.8498	0.8826	0.9031	0.9173	0.9278
9	0.6838	0.7635	0.8074	0.8365	0.8575	0.8736
11	0.6019	0.6878	0.7381	0.7730	0.7989	0.8193
13	0.5358	0.6233	0.6770	0.7153	0.7446	0.7660
15	0.4821	0.5687	0.6237	0.6640	0.6954	0.7209
17	0.4377	0.5222	0.5774	0.6185	0.6512	0.6781
19	0.4005	0.4823	0.5369	0.5783	0.6116	0.6394
21	0.3691	0.4479	0.5014	0.5388	0.5762	0.6043
25	0.3187	0.3915	0.4424	0.4824	0.5155	0.5438
31	0.2644	0.3280	0.3754	0.4128	0.4444	0.4718
35	0.2373	0.2972	0.3408	0.3763	0.4066	0.4332
41	0.2057	0.2594	0.2993	0.3321	0.3605	0.3856
61	0.1423	0.1821	0.2125	0.2383	0.2610	0.2816
101	0.0880	0.1140	0.1343	0.1520	0.1678	0.1825

$p=2,\ \alpha=0.99$				m			
n	7	8	9	10	12	15	20
3	0.9985	0.9987	0.9989	0.9990	0.9992	0.9993	0.9995
5	0.9774	0.9799	0.9819	0.9836	0.9861	0.9887	0.9914
7	0.9358	0.9423	0.9475	0.9519	0.9587	0.9660	0.9736
9	0.8862	0.8957	0.9051	0.9124	0.9239	0.9364	0.9501
11	0.8356	0.8493	0.8607	0.8705	0.8864	0.9039	0.9235
13	0.7872	0.8034	0.8171	0.8290	0.8486	0.8706	0.8958
15	0.7422	0.7602	0.7758	0.7894	0.8119	0.8378	0.8679
17	0.7008	0.7202	0.7374	0.7520	0.7770	0.8061	0.8404
19	0.6630	0.6834	0.7013	0.7172	0.7441	0.7757	0.8137
21	0.6285	0.0496	0.6682	0.6848	0.7132	0.7469	0.7880
25	0.5685	0.5902	0.6096	0.6271	0.6574	0.0694	0.7399
31	0.4962	0.5178	0.5374	0.5553	0.5868	0.6258	0.6760
35	0.4569	0.4782	0.4976	0.5154	0.5470	0.5867	0.6384
41	0.4082	0.4287	0.4475	0.4609	0.4961	0.5358	0.5887
61	0.3004	0.3179	0.3341	0.3494	0.3773	0.4142	0.4654
101	0.1961	0.2090	0.2211	0.2326	0.2542	0.2836	0.3264

（续表）

$p=3,\ \alpha=0.99$	m					
n	1	2	3	4	5	6
4	0.9933	0.9967	0.9978	0.9983	0.9987	0.9989
6	0.9260	0.9542	0.9664	0.9734	0.9779	0.9811
8	0.8302	0.8826	0.9086	0.9247	0.9359	0.9441
10	0.7400	0.8074	0.8434	0.8679	0.8552	0.8983
12	0.6628	0.7381	0.7816	0.8113	0.8334	0.8505
14	0.5981	0.6770	0.7245	0.7582	0.7837	0.8035
16	0.5439	0.6237	0.6735	0.7096	0.7376	0.2601
18	0.4981	0.5774	0.6282	0.6657	0.6953	0.7195
20	0.4591	0.5369	0.5879	0.6262	0.6568	0.6821
26	0.3711	0.4424	0.4914	0.5296	0.5610	0.5876
30	0.3286	0.3954	0.4423	0.4794	0.5103	0.5369
36	0.2804	0.3408	0.3842	0.4191	0.4746	0.4746
40	0.2553	0.3119	0.3531	0.3865	0.4151	0.4402
60	0.1763	0.2188	0.2509	0.2777	0.3012	0.3223
100	0.1088	0.1369	0.1586	0.1772	0.1939	0.2092

$p=3,\ \alpha=0.99$	m						
n	7	8	9	10	12	15	20
4	0.9990	0.9992	0.9993	0.9993	0.9994	0.9996	0.9997
6	0.9835	0.9853	0.9868	0.9880	0.9899	0.9918	0.9937
8	0.9504	0.9554	0.9595	0.9628	0.9682	0.9738	0.9797
10	0.9086	0.9170	0.9240	0.9298	0.9361	0.9492	0.9602
12	0.8643	0.8757	0.8853	0.8934	0.9066	0.9212	0.9374
14	0.8206	0.8344	0.8462	0.8564	0.8726	0.8918	0.9130
16	0.7788	0.7947	0.8083	0.8201	0.8397	0.8621	0.8879
18	0.7398	0.7571	0.7721	0.7853	0.8074	0.8329	0.8628
20	0.7035	0.7220	0.7381	0.7524	0.7764	0.8046	0.8381
26	0.6107	0.6309	0.6490	0.6651	0.6931	0.7267	0.7684
30	0.5601	0.5808	0.5993	0.6160	0.6452	0.6810	0.7263
36	0.4974	0.5180	0.5366	0.5536	0.5836	0.6211	0.6697
40	0.4626	0.4828	0.5012	0.5181	0.5482	0.5862	0.6360
60	0.3415	0.3592	0.3757	0.3910	0.4190	0.4555	0.5058
100	0.2234	0.2367	0.2492	0.2611	0.2832	0.3132	0.3564

（续表）

$p=4, \alpha=0.99$			m			
n	1	2	3	4	5	6
5	0.9950	0.9975	0.9983	0.9987	0.9990	0.9992
7	0.9411	0.9636	0.9734	0.9789	0.9825	0.9850
9	0.8591	0.9031	0.9247	0.9381	0.9473	0.9541
11	0.7779	0.8365	0.8679	0.8885	0.9032	0.9144
13	0.7057	0.7730	0.8113	0.8374	0.8567	0.8716
15	0.6434	0.7153	0.7582	0.7882	0.8110	0.8290
17	0.5902	0.6649	0.2096	0.7424	0.7677	0.7881
19	0.5440	0.6185	0.6657	0.7003	0.7275	0.7495
21	0.5044	0.5783	0.6262	0.6619	0.6903	0.7136
25	0.4395	0.5108	0.5586	0.5951	0.6247	[illegible]
31	0.3679	0.4337	0.4794	0.5152	0.5450	[illegible]
35	0.3316	0.3937	0.4375	0.4724	0.5016	0.5268
41	0.2888	0.3457	0.3865	0.4196	0.4476	0.4722
61	0.2016	0.2452	0.2777	0.3048	0.3284	0.3495
101	0.1255	0.1548	0.1772	0.1964	0.2135	0.2291

$p=4, \alpha=0.99$				m			
n	7	8	9	10	12	15	20
5	0.9993	0.9994	0.9994	0.9995	0.9996	0.9997	0.9997
7	0.9869	0.9884	0.9846	0.9905	0.9920	0.9935	0.9950
9	0.9593	0.9634	0.9668	0.9696	0.9739	0.9785	0.9834
11	0.9231	0.9302	0.9361	0.9410	0.9489	0.9574	0.9666
13	0.8836	0.8934	0.9017	0.9088	0.9202	0.9327	0.9466
15	0.8436	0.8559	0.8663	0.5752	0.8898	0.9062	0.9247
17	0.8049	0.8190	0.8312	0.5417	0.8592	0.8790	0.9019
19	0.7680	0.7837	0.7974	0.8092	0.8291	0.8519	0.8787
21	0.7333	0.7502	0.7650	0.7780	0.7801	0.8254	0.8557
25	0.7708	0.6894	0.7057	0.7203	0.7452	0.7748	0.8110
31	0.5926	0.6122	0.6297	0.6456	0.6731	0.7067	0.7489
35	0.5490	0.5688	0.5867	0.6029	0.6313	0.6664	0.7114
41	0.4939	0.5135	0.5314	0.5477	0.5767	0.6130	0.6605
61	0.3686	0.3862	0.4025	0.4177	0.4452	0.4810	0.5301
101	0.2435	0.2570	0.2696	0.2816	0.3039	0.3339	0.3771

（续表）

$p=5$, $\alpha=0.99$			m		
n	1	2	3	5	6
6	0.9960	0.9980	0.9987	0.9990	0.9992
8	0.9510	0.9699	0.9779	0.9825	0.9855
10	0.8793	0.9173	0.9359	0.9473	0.9552
12	0.8056	0.8575	0.8552	0.9032	0.9161
14	0.7381	0.7989	0.8334	0.8567	0.8738
16	0.6783	0.7446	0.7837	0.8110	0.8315
18	0.6264	0.6954	0.7376	0.7677	0.7909
20	0.5810	0.6512	0.6953	0.7275	0.7525
26	[illegible]754	0.5443	0.5900	0.6247	0.6528
30	[illegible]34	0.4895	0.5345	0.5694	0.5980
36	[illegible].3633	0.4250	0.4676	0.5016	0.5300
40	0.3317	0.3900	0.4313	0.4644	0.4922
60	0.2309	0.2763	0.3098	0.3374	0.3614
100	0.1434	0.1741	0.1975	0.2173	0.2349

*) 此表中 θ_{max}是 $|A-\theta(A+B)|=0$ 的最大特征根，其中 $A\sim W_p(m,\Sigma)$，$B\sim W_p(n,\Sigma)$ ($\Sigma>0$ 任意)